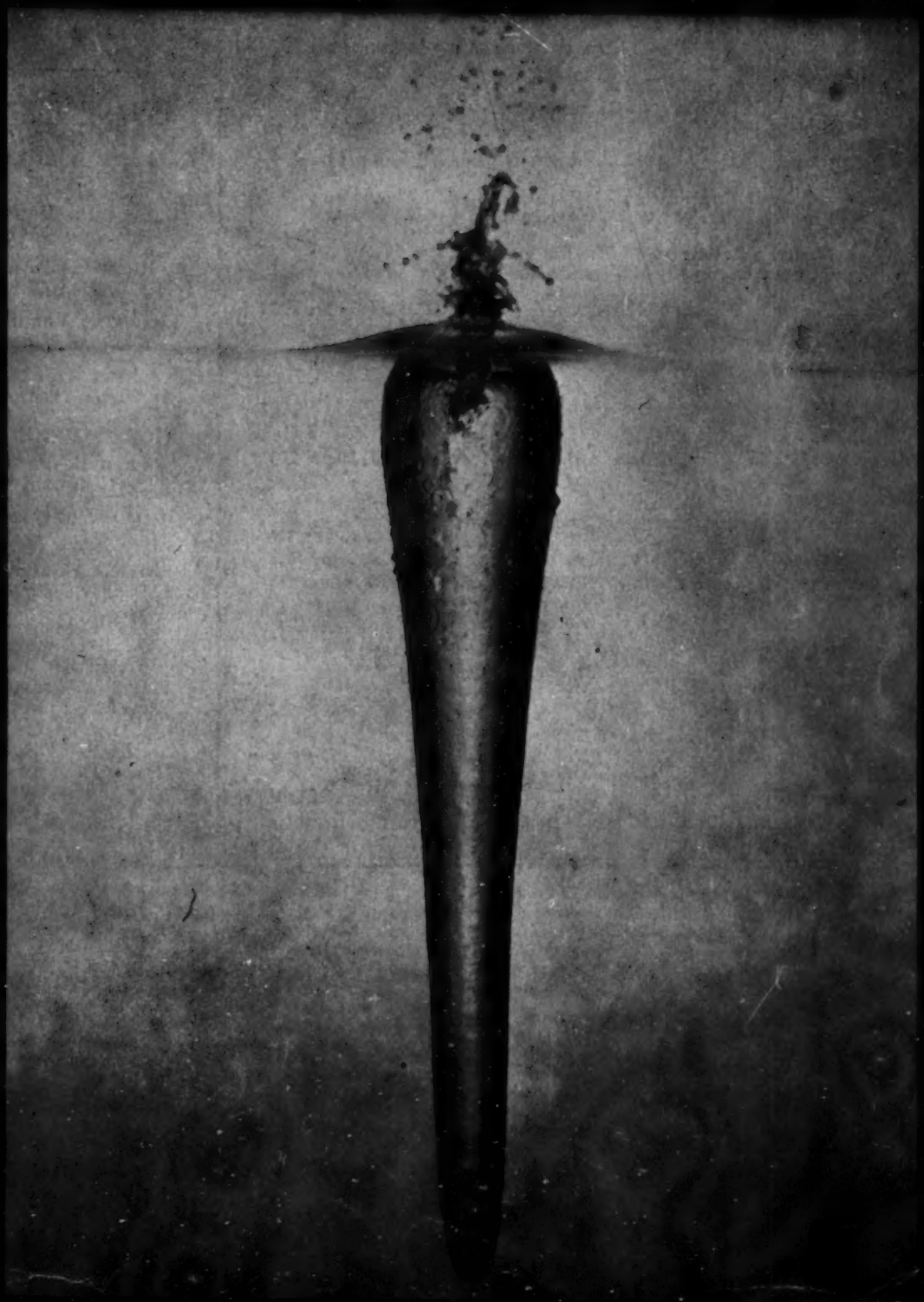
JOURNAL OF APPLIED PHYSICS



JOURNAL
OF
APPLIED PHYSICS

Vol. 19, No. 2 FEBRUARY, 1948

In This Issue

Page

Influence of Atmosphere Pressure on the Phenomena Accompanying the Entry of Spheres into Water DAVID GILBARG AND ROBERT A. ANDERSON 127

Statistical Aspects of Fracture Problems BENJAMIN EPSTEIN 140

Electronic Differential Analyzer J. S. KOEHLER 148

Rising Sun Magnetron S. MILLMAN AND A. T. NORDSIECK 156

Resonant Modes of the Rising Sun and Other Unstrapped Magnetron Anode Blocks NORMAN M. KROLL AND WILLIS E. LAMB, JR. 166

Palletron, a New Electron Resonator and Its Proposed Application to the Generation of Potentials in the Million-Volt Range A. M. SKELLETT 187

Measurement of Stress by Means of X-Rays D. E. THOMAS 190

Measured Impedances of Helical Beam Antennas OTTO J. GLASSER AND JOHN D. KRAUS 193

Dark-Field Electron Microscopy. I. Studies of Crystalline Substances in Dark-Field C. E. HALL 198

Tensile Properties of Rolled Silver Chloride B. M. AXILROD AND J. J. LAMB 213

Letters to the Editor:

Note on the Anomalous Physical Properties of Processed Atmospheric Nitrogen CLARKE C. MINTER 217

Self-Diffusion in Iron C. ERNEST BIRCHENALL AND ROBERT F. MEHL 217

Temperature Gradients for Convection in Well Models MARY JANE AULD 218

In Every Issue

Calendar of Meetings 216

New Books 218

New Booklets 219

Here and There 219

Index to Advertisers XXXIV

BOARD OF EDITORS

ELMER HUTCHISSON,
Editor

Associate Editors

R. BOWLING BARNES
OTTO BEECK
W. F. BUSSE
W. JAMES LYONS
PHILIP MORRISON
PAUL E. SABINE
FREDERICK SEITZ
J. R. TOWNSEND
JOHN VON NEUMANN

Publication of the
American Institute of Physics

ADMINISTRATIVE STAFF

HENRY A. BARTON,
Director

CLEVELAND NORCROSS,
Assistant Director

RUTH F. BRYANS,
Publication Manager

THEODORE VORBURGER,
Business Manager

JOHN T. TATE,
Adviser on Publications

GOVERNING BOARD

GEORGE R. HARRISON,
Chairman

WALLACE WATERFALL,
Secretary

GEORGE B. PEGRAM,
Treasurer

E. U. CONDON
A. J. DEMPSTER
W. F. FAIR, JR.
FLOYD A. FIRESTONE
HARVEY FLETCHER
R. C. GIBBS
K. S. GIBSON
PAUL E. KLOPSTEG
A. H. PFUND
JOHN T. TATE
M. W. WHITE
A. G. WORTHING

The JOURNAL OF APPLIED PHYSICS, published monthly at Prince and Lemon Sts., Lancaster, Pa., is devoted to physics in its role as the science basic to other natural sciences and to the arts and industries. Previous to 1937 this publication was known as PHYSICS. The JOURNAL OF APPLIED PHYSICS publishes editorials, news of physicists, and reviews, as well as technical papers of applied physics.

Manuscripts should be submitted to Elmer Hutchisson, Editor, Case Institute of Technology, Cleveland 6, Ohio.

Subscriptions and orders for back numbers should be addressed to Prince and Lemon Streets, Lancaster, Pa. or to the American Institute of Physics, 57 East 55 Street, New York 22, N. Y.

Subscription Price

	U.S. AND CANADA	ELSEWHERE
To members and associate members of the American Institute of Physics	$6.00	$7.00
To all others	8.00	9.00

Back Number Prices

Complete set: Vol. 1, July, 1931–Vol. 18, 1948—$132.00
Yearly back number rate: $10.00
Single copies: $1.00 each.

Changes of address should be sent to the Business Manager, American Institute of Physics; **corrected proofs** should be sent to the Publication Manager, American Institute of Physics.

Advertising rates supplied on request. Orders, advertising copy, and cuts should be sent to the American Institute of Physics.

Entered as second class matter January 22, 1937, at the Post Office at Lancaster, Pennsylvania, under the Act of March 3, 1879. Accepted for mailing at the special rate of postage provided for in the Act of February 28, 1925, authorized May 2, 1932.

Copyright 1948, by the American Institute of Physics.

Journal
of
Applied Physics

Volume 19, Number 2 February, 1948

Influence of Atmospheric Pressure on the Phenomena Accompanying the Entry of Spheres into Water

DAVID GILBARG AND ROBERT A. ANDERSON
Naval Ordnance Laboratory, Washington, D. C.

(Received January 23, 1947)

High speed motion-picture studies were made of the dependence of air-water entry cavities on the relevant physical parameters—in particular, on the atmospheric pressure over the water surface. Experiments on the vertical entry of spheres ¼ in. to 1 in. in diameter, with entrance velocities between 15 ft./sec. and 100 ft./sec., and air pressures between 1/60 atmosphere and 3 atmospheres, show that surface sealing of the cavity is a major factor in controlling cavity formation, and is the factor most responsible for non-Froude scaling of cavity phenomena. Surface sealing is a function chiefly of the atmospheric density and projectile velocity, although surface tension is also important in its effect on splash formation. Froude's scaling is found to hold true in the region of low Froude's numbers and low atmospheric pressures. The measured average drag coefficient of the projectiles is independent of cavity size and shape. Jets are found to occur regularly and with great strength at both deep and surface closures of the cavity. The observed presence of jets in finite cavities is predicted by an extension of the Kirchhoff-Helmholtz theory of the infinite (two-dimensional) cavity to the case of finite cavities with finite cavitation numbers. Some of the results of this theory are mentioned.

INTRODUCTION

THE phenomena accompanying the air-water entry of projectiles have been studied qualitatively to some extent by previous investigators[1,2]; Davies[2] has shown in a series of interesting experiments that the cavity and splash phenomena accompanying water entry are considerably influenced by the air pressure above the water surface, that is, the atmospheric pressure, but as yet a systematic account of the influence of atmospheric pressure on these phenomena, and its effect on the scaling of water-entry phenomena, has not been given. We shall describe here some of the results obtained thus far at the Naval Ordnance Laboratory as part of a program to explore water-entry problems.

When a projectile, say a sphere, is shot vertically into water from air, the phenomena of interest are the splash, cavities, and jets. For purposes of this general discussion, we shall assume projectile diameters between ¼ in. and 1 in., and velocities between 15 and 100 ft./sec. From the instant of contact a thin sheet of water travels up the surface of the projectile and then detaches itself from the surface. Worthington's[1] investigations have shown that for sufficiently low entrance velocities the sheet does not detach itself, but closes about the surface. After detachment this sheet, the "splash," rises symmetrically and, depending on the entrance conditions (such as projectile velocity, atmospheric density, etc.) may close to form a domed closed surface excluding the passage of air into the air cavity

[1] A. M. Worthington, *A Study of Splashes* (Longmans Green and Company, New York, 1908).
[2] R. M. Davies, "Influence of atmospheric pressure on the phenomena accompanying the fall of small scale projectiles into a liquid" (Engineering Laboratory, Cambridge, England, 1944).

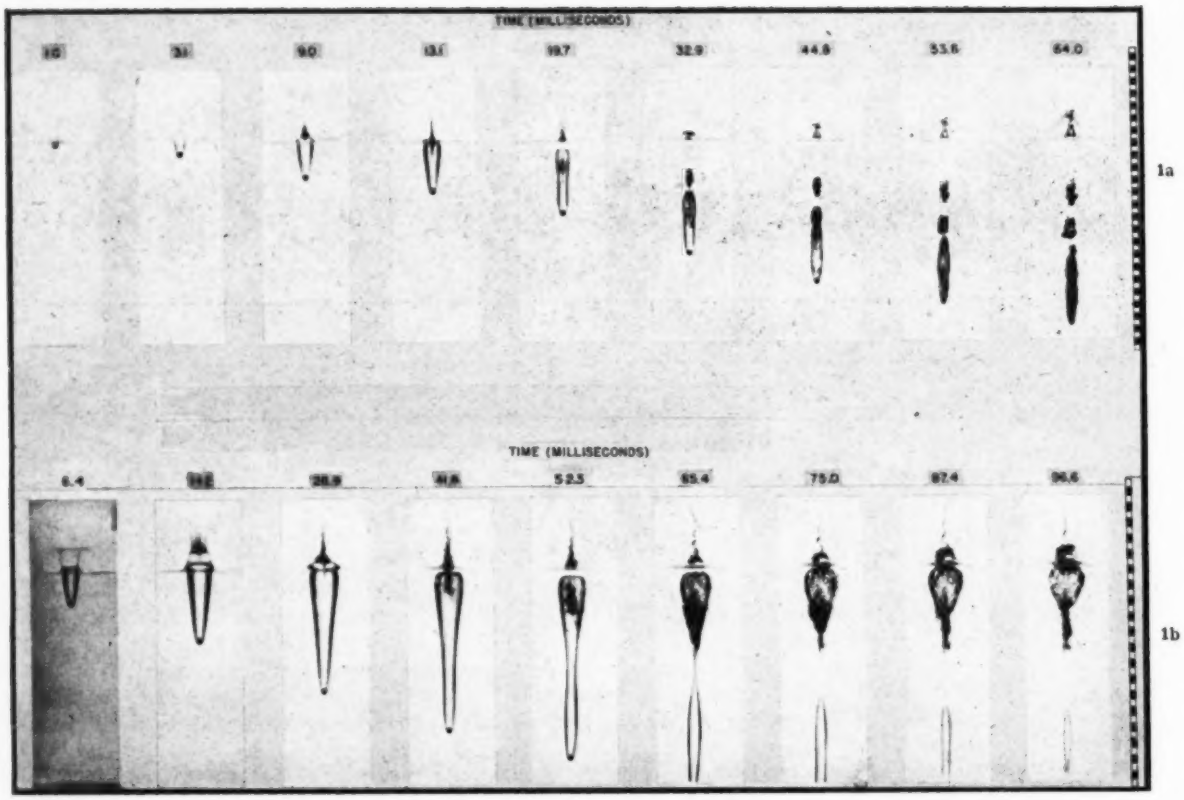


FIG. 1a. Vertical entry of 1 in. steel sphere at 70 ft./sec. and normal atmospheric pressure. Early surface closure prevents further growth of cavity as well as the occurrence if a distinct deep closure. The down-jet from surface closure is reinforced by the thickening of the dome forming the surface seal and travels down the cavity to strike the model forcefully.

FIG. 1b. Vertical entry of 1 in. steel sphere at 70 ft./sec. and $\frac{1}{3}$ atmospheric pressure, shows entry with surface closure as well as distinct deep closure. A long closed cavity or bubble trails the model after deep closure.

beneath it; Figs. 1a and 1b show entry with closed splash, Fig. 1c without closure.

As the projectile penetrates the medium it leaves behind it an air cavity, which first grows because of the energy imparted to the water by the moving body, and then contracts and closes at some point between the body and the surface, thus separating the cavity into two parts: a cavity connected to the surface, and a closed cavity, sometimes called a bubble, surrounding the projectile as shown in Fig. 1b. The cavity can thus seal at both the surface, and at a subsurface point. We shall refer to these closures as "surface closure" and "deep closure," respectively.

Accompanying all closures are jets, which can be understood as the result of collisions between converging bodies of water in the presence of a free boundary. Upward jets are always present at deep closure, starting first as a thin needle-like stream, traveling upward from the closure point, and increasing in size as the cavity boundary rises towards the surface, as shown in Fig. 1c. Such jets are known from Worthington's studies. The velocity of this up-jet is initially somewhat greater than that of the projectile. At both surface and deep closures down-jets also are produced which travel down the cavity and, in many observed cases, strike the projectile at the extreme lower end, as in Fig. 1a. According to the theory of the finite cavity, jets are inherent in steady flows with finite cavities.

SPLASH AND SURFACE CLOSURE

As the splash rises from the surface it is subject to two chief forces leading to its collapse:

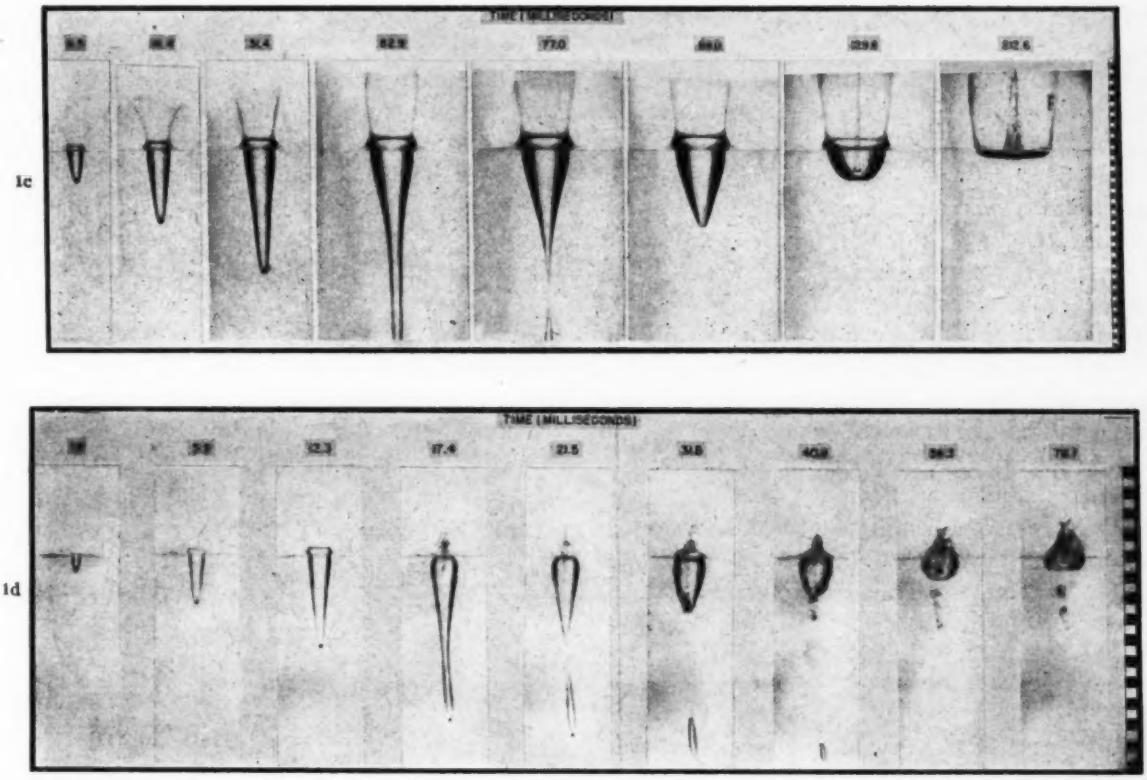


FIG. 1c. Vertical entry of 1 in. steel sphere at 70 ft./sec. and $\frac{1}{16}$ atmospheric pressure. Splash remains open at reduced pressure because of absence of Bernoulli effect. Cavity expands laterally because surface closure does not occur. The development of the up-jet from deep closure is shown.

FIG. 1d. Vertical entry of $\frac{1}{4}$ in. steel sphere at 35 ft./sec. and $\frac{1}{16}$ atmospheric pressure. Shot has been scaled according to Froude's law with shot in Fig. 1c. Dissimilarity of the two cavities shows failure of Froude scaling.

(1) an under-pressure caused by the flow of air into the cavity behind the projectile, and (2), surface tension. The underpressure causing closure may be found by treating the splash as a hollow cylinder into which air is rushing with a velocity equal to that of the projectile; assuming potential flow, the pressure on the sheath is essentially equal to $\frac{1}{2}\rho_A V^2$, where ρ_A is the density of the gas and V the velocity of the projectile. The effect of increased atmospheric density and entrance velocity of the sphere is to increase the Bernoulli pressure, $\frac{1}{2}\rho_A V^2$, and thus to diminish the time of surface closure. It should be noted that although the active agent in surface closure is the air *density*, rather than its pressure, we shall hereafter refer to the *pressure* above the water surface as long as there is no ambiguity as to the value of the density. Therefore, unless otherwise stated, when a certain fraction of atmospheric pressure is mentioned, the same fraction of the density is implied.

For pressures of one atmosphere or greater the splash closes very quickly behind the projectile, as in Fig. 1a, whereas for lower pressures, say 1/16 atmosphere as in Fig. 1c, the splash remains open throughout. The pressures for which surface closure does not occur are dependent on the scale.

The typical dependence of surface closure on atmospheric pressure and velocity is shown in Figs. 2a, 2b, and 2c for the three sizes of spheres studied. Both time and velocity are scaled up to 1-in. prototype by Froude's law, i.e., time $\propto \lambda^{\frac{1}{2}}$, velocity $\propto \lambda^{\frac{1}{2}}$, where λ = scale factor. This practice is followed in all the curves shown. To a first approximation the data show that for a fixed air pressure the time of surface closure is inversely proportional to entrance velocity. The

quantitative dependence of surface closure on pressure does not seem to follow a simple mathematical law. For the pressures tested lower than those shown on the curves, surface closure did not occur at all.

Table I lists some specific data concerning splash dimensions under various entry conditions. It should be noted that under the same atmospheric conditions, and scaled entrance velocities, the splash height is out of proportion with the scale of the projectile, the splash for the 1-in. sphere being proportionately much higher than for the smaller projectiles under corresponding conditions. As a check on the possible effect of atmospheric pressures on the splash, as distinguished from the density, several experiments were conducted with a freon-air mixture as atmosphere, the pressures in all cases being a fraction of normal atmospheric pressure, while the density was that of the normal atmosphere. In all of these cases the splash behavior was identical with that in the experiments having both normal density and pressure.

The results given in the preceding paragraph would seem to imply, in the absence of any other relevant parameters (except possibly viscosity), that surface tension is the major factor in causing non-similarity of splash phenomena for different projectile sizes.

Although surface tension is undoubtedly important in splash formation, its influence on surface closure is probably minor for all but low velocities and small projectiles, as the following estimates show. Assuming the splash is cylindrical at its maximum growth, the surface-tension pressure tending to collapse the splash is $2T/r$, where T is the surface tension and r = radius of the cylinder. The relative magnitude of surface

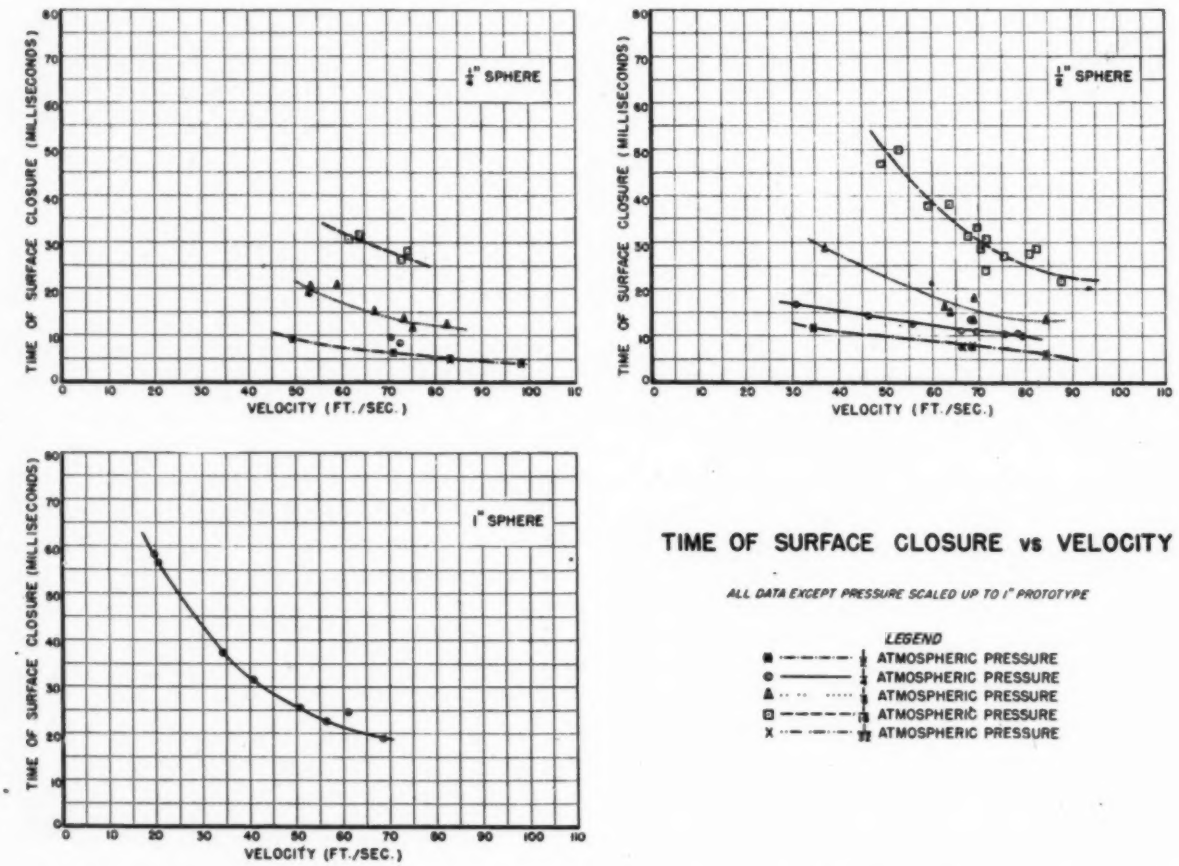
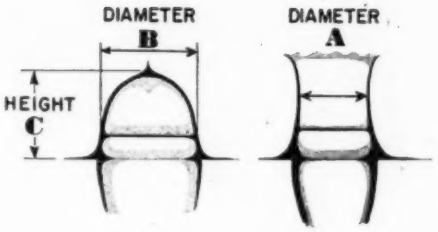


FIG. 2a, b, c. Time of surface closure *vs.* velocity, all data except pressure scaled up to 1" prototype according to Froude's law. Closure time falls off with increase of sphere velocity. Note that closure times for different size models do not coincide when Froude scaling alone is used.

TABLE I. Splash dimensions for various entry conditions.



Model Size	Pressure (Atmosphere)	Velocity ft/sec.	Diameter of Cylindrical Splash in Model Diameters A	Diameter of Splash at Surface Closure in Model Diameters B	Height of Splash at Surface Closure in Model Diameters C
1/4	1/4	35	2.6	2.8	2.6
1/4	1/8	34	3.2	3.6	3.6
1/4	1/16	36	6.0	4.8	5.6
1/2	1/4	49	2.8	3.0	2.6
1/2	1/8	49	3.6	4.0	5.6
1/2	1/16	50	4.6	5.0	8.0
1	1/4	69	3.0	3.6	3.4
1	1/8	72	4.3	4.8	5.7
1	1/16	69	8.0	No S. C.	No S. C.

tension and underpressure forces causing closure is of the order

$$(2T/r)/(\tfrac{1}{2}\rho_A V^2) = (4T)/(r\rho_A V^2) = N;$$

thus the greater scale, atmospheric density, and velocity, the smaller the relative contribution of surface tension towards surface closure. For a 1/4-in. sphere at 1/2 atmosphere and velocity 36 ft./sec. $N = 0.32$; for a 1-in. sphere at 1/16 atmosphere at velocity 69 ft./sec. $N = 0.005$. As this shows, N is quite small except for small spheres at higher pressures and lower velocities. It would appear from these values that surface closure largely is the result of the underpressure caused by air flow into the cavity. As the dome starts to close over, and its curvature increases, the effect of surface tension is presumably greater, and would tend to speed up the closure. Indeed, closing over of the dome is found to be very rapid once it has begun.

The completely surface character of the splash, namely, the fact that it is derived from the surface water layer, was revealed in photographs which showed the motion of surface particles near the point of entry going into and up the splash during its formation. The sensitivity of the splash to disturbed surface conditions was demonstrated in a number of cases where the surface was almost inperceptibly ruffled by droplets of water preceding the entry of the projectile. In these cases the splash was badly malformed, in contrast with its usual symmetry, and did not close under conditions for which it would normally.

CAVITY AND DEEP CLOSURE

As long as the cavity is open to the atmosphere, the external pressure has little importance in cavity formation. For the motion of the water is given by the potential equation, subject to the boundary condition represented by Bernoulli's equation

$$(p - p_A)/\rho = gz - \tfrac{1}{2}V^2 - \partial\phi/\partial t,$$

where ϕ = velocity potential, i.e., $\partial\phi/\partial x = v_x$ etc., v = flow velocity, g = gravitational constant = 32 ft./sec.^{2}, z = depth below water surface, p = pressure in the water, p_A = atmospheric pressure over the water surface, and ρ = water density. On the boundary, $p = p_A$ approximately,[3] so that the atmospheric pressure does not appear in the equations governing cavity formation.

With surface closure the further passage of air into the cavity ceases, so that the normal continued expansion of the cavity is inhibited by the resulting underpressure in the cavity. The result is smaller cavity size and earlier deep closure, as in Fig. 3. Indeed, for sufficiently early surface closure, deep closure is found not to occur at all (see Fig. 1a, for example). The dependence of deep closure on surface closure is shown clearly in the curve of Fig. 4. It should be noted that, within experimental error, the time of deep closure, T_d, is a function only of the time of surface closure, T_s, since all points fall on a single curve, irrespective of velocity, scale, and pressure. That this should be the case can be inferred rigorously as follows: If the times of surface closure and of deep closure are dependent only on the impact velocity, V, reference dimension l, air pressure p, and air density ρ, then by

[3] This equality, $p = p_A$, holds except for an underpressure of the order $\frac{1}{2}\rho_A V^2$ (V = projectile velocity), which is considerably smaller than p_A at the velocities we are considering.

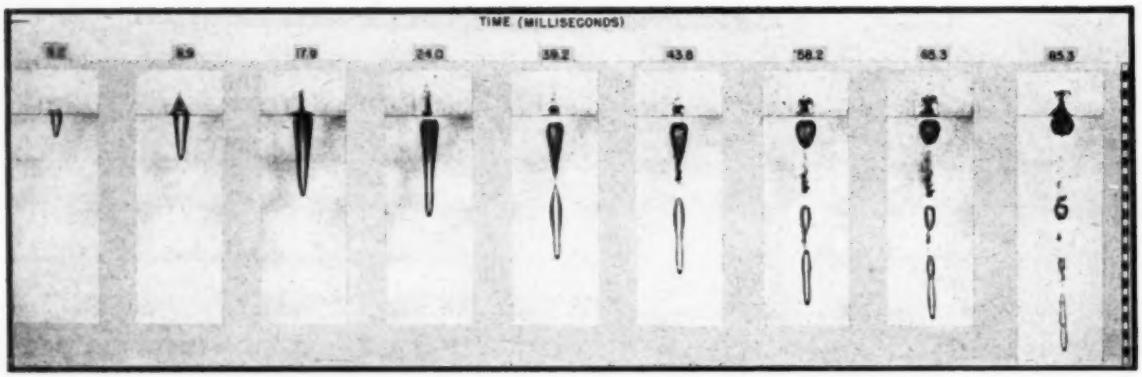


FIG. 3. Vertical entry of ½" steel sphere at 50 ft./sec. and ¼ atmospheric pressure. Shows small cavity size and early deep closure as result of early surface closure.

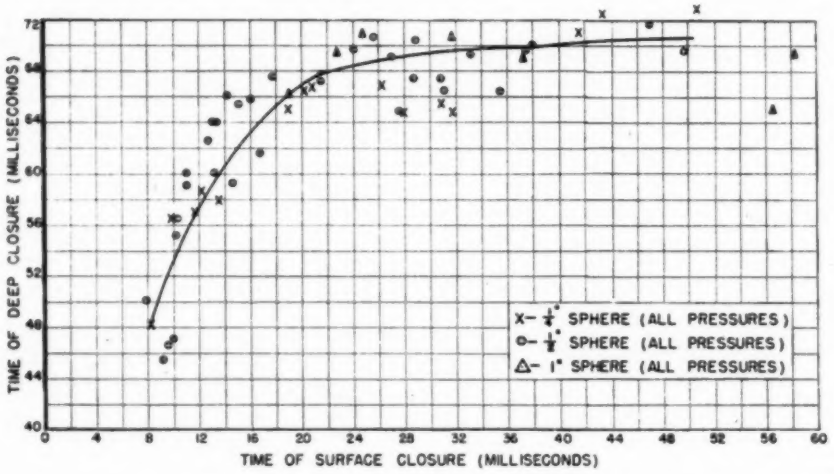

TIME OF DEEP CLOSURE (MILLISECONDS)
TIME OF SURFACE CLOSURE (MILLISECONDS)
X- ½" SPHERE (ALL PRESSURES)
o- ¼" SPHERE (ALL PRESSURES)
Δ- 1" SPHERE (ALL PRESSURES)

FIG. 4. Curve of time of deep closure *vs.* time of surface closure. (All times scaled, according to Froude's law, up to 1" prototype.) Illustrating that time of deep closure for early closures is a function of time of surface closure only. For late closure the time of deep closure remains about 70 milliseconds.

dimensional analysis one can write

$$T_s V/l = f_1(V^2/lg,\ p/p_0,\ \rho/\rho_0),$$
$$T_d V/l = f_2(V^2/lg,\ p/p_0,\ \rho/\rho_0),$$

where p_0 and ρ_0 are a reference pressure and density, respectively, such as that of the standard atmosphere. The dependence on surface tension can be omitted since it is a constant throughout. Now, since the pressure above the water surface is reduced isothermally in these experiments, $\rho/\rho_0 = p/p_0$, so that explicit reference to the density can be removed from each of the above equations. The pressure can then be eliminated between the resulting two equations to give a simple relation of the form

$$T_d V/l = f(T_s V/l,\ V^2/lg).$$

This is precisely the statement that the time of deep closure is a function only of the time of surface closure, provided both times are scaled up to a single prototype dimension by Froude's law, as in Fig. 4. The curves show also that, after a certain time of surface closure, $T_s \approx 20$ millisec., the time of deep closure is essentially constant $T_d \approx 70$ millisec.

The variation of depth of deep closure with atmospheric pressure and entrance velocity is shown in Figs. 5a–5c. These curves show in another way the dependence of the cavity on events at the surface. It was found that the time of deep closure is essentially a constant independent of entrance velocity for low pressures and hence for late surface closures. However, for

higher pressures, i.e., early surface closure, the time of deep closure is somewhat reduced, and for the same reason is a decreasing function of velocity. Figures 5a, 5b, and 5c show corresponding effects for depth of deep closure. These results show, in other words, that atmospheric pressure is of importance in cavity formation only indirectly as it affects surface closure, which in turn affects cavity growth. The inhibiting effect of surface closure on cavity size is shown in Figs. 1a, 1b, and 1c, for all of which the entrance velocity is the same, but with surface closures occurring at different times. An interesting example is shown by Figs. 6a and 6b, for which the entry conditions were identical in every respect except that the malformation of one splash prevented surface closure, whereas the other closure was normal. The difference in cavity diameters should be noted.

JETS

Quantitative information concerning jet phenomena was difficult to obtain for experimental reasons. However, the following qualitative observations were made. The up-jet at deep closure, as Fig. 1c shows, starts upwards at high speed as a thin stream; velocities 1.5 times the projectile velocity were measured in several instances. The true maximum jet velocities are probably greater than those measured. As the cavity rises to the surface the jet broadens in size, and its base forms the flat bottom of the cavity seen in Figs. 1c and 6b. The up-jets are most powerful in the absence of surface closure, and become less so as surface closure occurs earlier. Thus, at the lowest atmospheric pressures, the up-jets are strongest.

The up-jet can be understood physically as the result of water masses colliding at deep closure

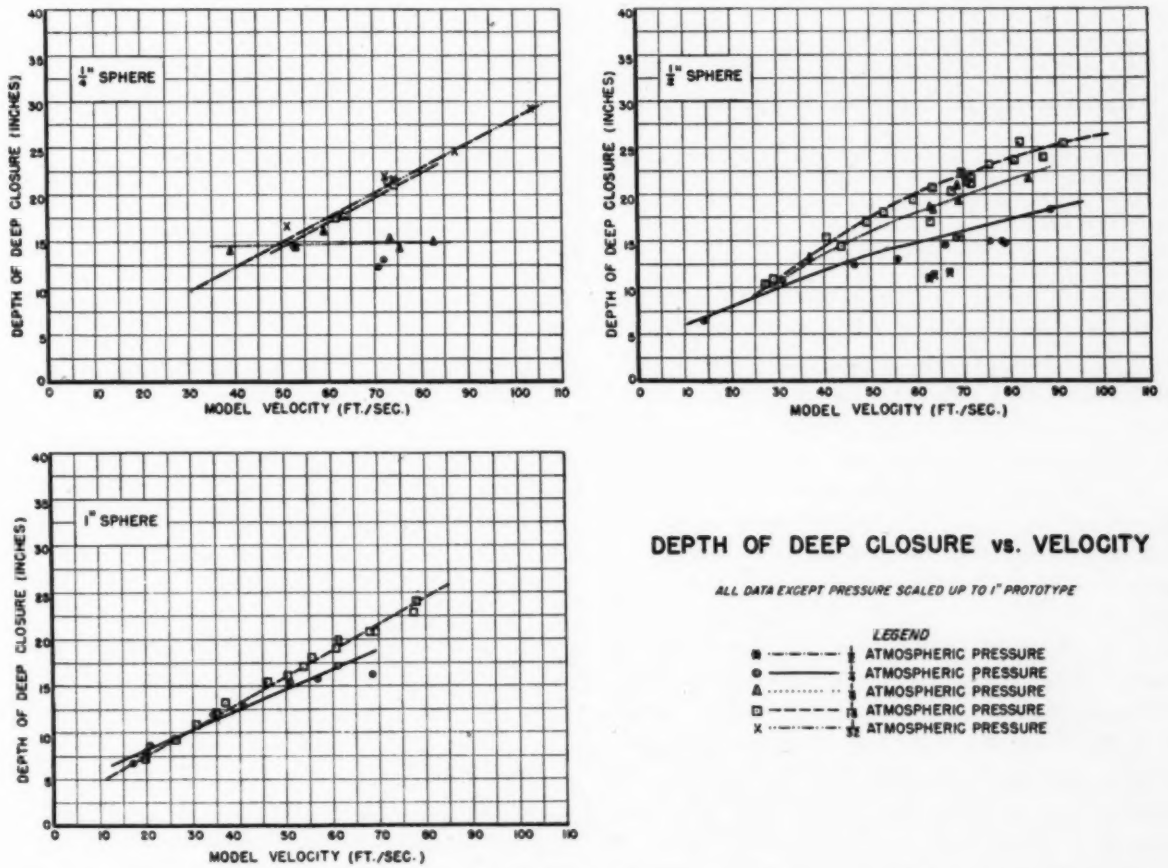


FIG. 5. Curves of depth of deep closure *vs.* velocity. (All data except pressure scaled up to 1″ prototype according to Froude's Law.) For the lower pressures, where the surface closures do not inhibit normal cavity development, the curves are essentially the same.

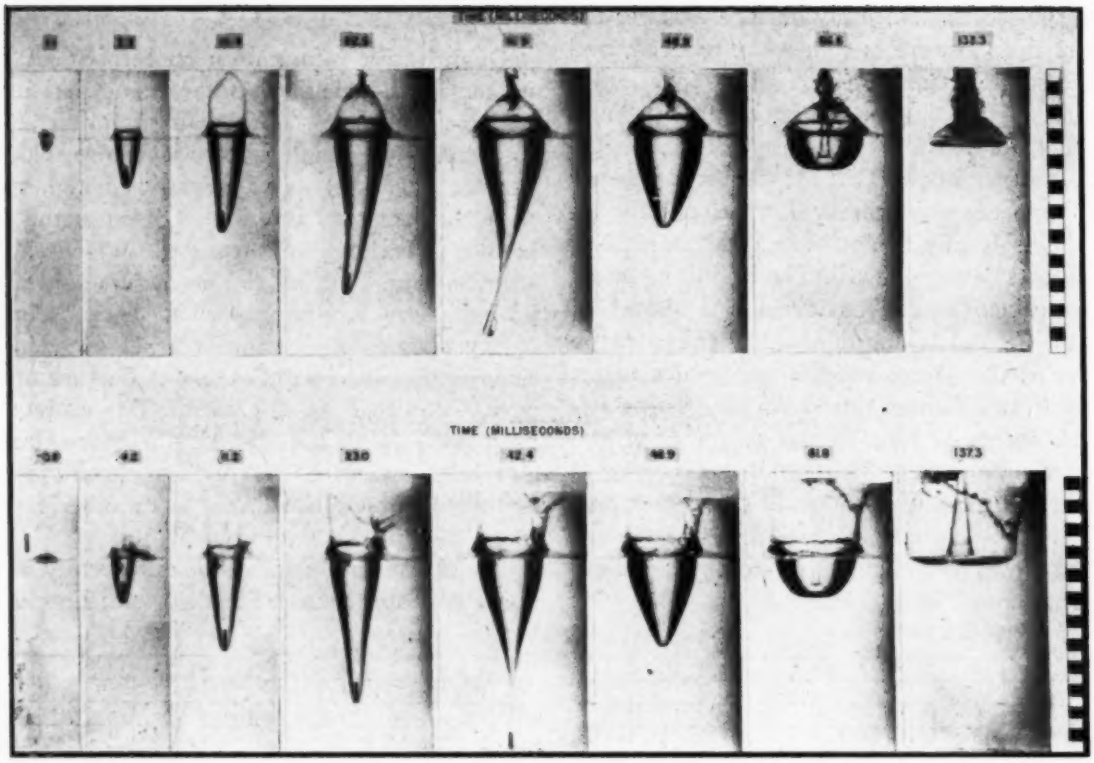


FIG. 6a, b. Vertical entry of $\frac{1}{4}$-in. Duralumin cylinder at 50 ft./sec. and $\frac{1}{32}$ atmospheric pressure. Example of entry under conditions identical in every respect except that malformation of one splash prevented surface closure, whereas the other closure was normal. In Fig. 6b the ruffled water surface is shown just before the entry of the model. Note the difference in cavity diameters.

because of the restoring action of hydrostatic pressure. This is at least partly confirmed by the fact that even for non-vertical entries, the up-jet rises vertically, although the cavity axis is not vertical. For oblique, low angle entry, up-jets have no opportunity to form for this reason.

Down-jets occur at every closure. The down-jets following deep closure, and particularly for low pressures, can be strong, and have been observed to deflect the projectile's course, or to deform the cavity wall. At higher pressures, 1/4 to 1/8 atmospheres depending on the scale, the deep closure is somewhat weakened, and the subsequent down-jets, although visible, are not as strong, nor do they overtake the projectile. The reverse situation prevails with regard to jets produced at surface closure. At low pressures, e.g., 1-in. sphere, 1/8 atmosphere, 70 ft./sec., the down-jet produced with the closure of the thin-walled dome is weak, whereas at the higher pressures, e.g., 1-in. sphere, 1/2 atmosphere, 70 ft./sec. the jet is initially stronger because of the more forceful closure and, in addition, the jet is reinforced by the thickening of the dome as it is drawn below the water surface; this is seen in Figs. 1a and 1b. In the latter cases the jet is readily observed traveling the full length of the cavity to strike the projectile.

Of particular interest in showing the basic nature of the closed cavity is the presence of a series of down-jets originating at the rear of the closed cavity without further closures taking place. It is true that the jets in these instances are considerably weaker than those which arise after closure, probably because of the turbulent dissipation at the rear of the cavity bubble, but they nevertheless confirm the theory of the finite cavity, the essential feature of which is the jet.

If steady flow with a cavity is considered, in the absence of hydrostatic forces, it is at once evident from Bernoulli's equation that the boundary of discontinuity must be a surface of

constant stream velocity relative to the projectile. In the case of a finite cavity the rear point of the cavity would be a stagnation point, in contradiction to the boundary condition; or, viewed physically, being a stagnation point, it would be a point of high pressure, which, in the presence of a free boundary, would give rise to jets into the cavity. The reverse jet is thus seen to be not only reasonable, but a physical necessity. However, as mentioned above, turbulent mixing at the rear of the cavity would tend to weaken the jet considerably. The significance of this picture of the finite cavity from the point of view of cavity growth and decay is discussed in a later paragraph.

SCALING

On the assumption that water-entry phenomena are functions only of inertial elements, pressures, gravity, and surface tension, to the exclusion of viscosity and compressibility, the usual dimensional analysis states that any non-dimensional quantity is a function only of the parameters: V^2/lg, $p_A/(\frac{1}{2}\rho_A V^2)$, ρ_A/ρ_0, $T/(\rho_A V^2)$ where all terms are as previously defined, and ρ_A = atmospheric density; l = some reference dimension, e.g., projectile diameter. The first term, V^2/lg, is the well-known Froude number, the next two represent the influence of atmospheric pressure and density, respectively, and the final term that of surface tension.

The condition for similarity in scaled experiments is that the four parameters be constant: that is, (1) $V \propto (l)^{\frac{1}{2}}$, (2) $p_A \propto l$, (3) ρ_A = constant, (4) $T \propto l^2$. Thus, scaling requires, in general, the latter three conditions on the atmosphere and surface tension, in addition to Froude's law, (1). Conditions (2) and (3) necessitate experimentally the use of dense gases, by means of which the atmospheric density can be kept constant, while the pressure is reduced.[4] However, condition (4) is a difficult one to meet experimentally and it is likely, as indicated in earlier paragraphs, that this will be responsible for non-similarity over a large region of entry conditions, however well requirements (1)–(3) are met.

As mentioned above, Froude's scaling applies as long as the effects of the atmosphere and surface tension are negligible. As was shown on page 132, the cavity-formation scales by Froude's law as long as the cavity is open to the atmosphere, for in this case the potential equation and the boundary condition Bernoulli's equation, scale according to Froude's law, independently of the atmosphere. One might correctly anticipate, therefore, that for low pressures, low velocities, or low Froude's numbers, for which surface closure is either late or does not occur, Froude scaling adequately describes the cavity. If Figs. 5a, 5b, and 5c are examined, it will be noted that the curves of depth of deep closure for the three different scale projectiles are essentially the same at the lower pressures. The maximum value of Froude's number for which similarity obtains depends on the pressure—the lower the pressure, the higher Froude's number. However, the range

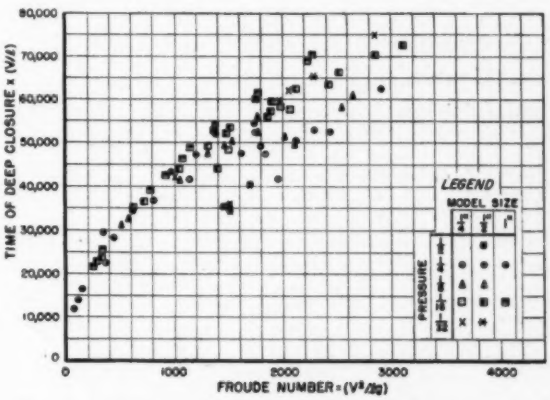


FIG. 7. Non-dimensional time of deep closure as a function of Froude number for various model sizes and atmospheric conditions.

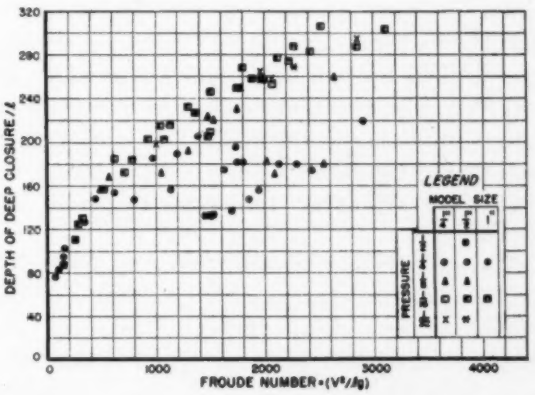


FIG. 8. Non-dimensional depth of deep closure as a function of Froude number for various model sizes and atmospheric conditions.

[4] The need for the use of dense gases was first pointed out by Professor Garrett Birkhoff.

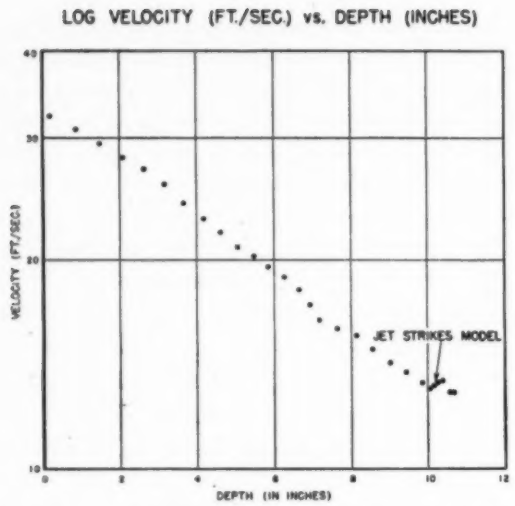


FIG. 9. Sample semi-log plot of velocity *vs.* depth for vertical entry of 1/4" diameter steel sphere at 36 ft./sec. and 1/34 normal atmospheric pressure. Effect on Model of impact of down-jet from deep closure is shown.

of applicability of Froude's scaling is not clearly determinable from the experimental data given in Figs. 7 and 8, which show the non-dimensional time and depth of deep closure as functions of Froude's number for the various atmospheric conditions.

The same boundary condition for the closed cavity, with cavity pressure unequal to atmospheric pressure, therefore shows that pressure scaling, $p \propto l$, is necessary for similarity of closed cavities. In the absence of gravity the condition for similarity in cavitation is equality of the cavitation parameter $(P-p)/(\frac{1}{2}\rho_W V^2)$, where P, p, are the static pressures in the free stream and cavity, respectively, and V is the velocity of the model. In the presence of gravity, with the Froude condition, $V^2 \propto l$, also imposed, it follows that $p \propto l$ is the condition on the pressure for similarity. The effect of non-scaling of the atmospheric pressure is to make the cavity proportionately too small or large, accordingly as the pressure is relatively too large or small, respectively.

The requirement that the atmospheric density, ρ_A, be constant for scaled experiments, can be seen to arise from the condition that the underpressure $\frac{1}{2}\rho_A V^2$ be proportional to the scale. As we have seen from the experiments discussed here, too low a density results in late surface closures and therefore in proportionately large cavities. The use of dense gases, by which both pressure and density are correctly scaled, has been attempted thus far in only a few cases—too few to present definitive quantitative information. However, in agreement with the above account the several experiments with a freon-air mixture, where the densities are equal, showed that the splash and surface closure are indeed the same, because of equal density, but the closed cavity was slightly wider for the lower pressure.

The effect of surface tension on similarity of cavity phenomena is through the medium of the splash, and is of importance insofar as surface closure influences the scaling of other cavity phenomena. As pointed out previously, surface tension forces are only secondary in causing surface closure, but are probably of primary importance in determining splash formation. As a previous paragraph shows, and also a comparison of Figs. 1c and 1d, the splash, under Froude scaling, does not scale. This is also revealed by comparison of Figs. 2a, 2b, and 2c; the curves of time of surface closure for the different sphere sizes should coincide when the external pressures are the same, but actually they do not. Thus, indirectly, through the medium of splash formation, surface tension is responsible for non-scaling of cavity phenomena, even when the atmospheric pressure and density are controlled. This would recommend the use of larger size models in water-entry model studies.

DRAG

The effect of the highly varied cavity shapes on the drag was studied for 1/4-in. steel spheres by analysis of the high speed motion pictures. Assuming a constant drag coefficient, C_D, for the projectile, and entrance velocities sufficiently high that cavity can be neglected over the studied portion of underwater path,[5] then the equation of motion of the projectile is

$$m dV/dt = -C_D \tfrac{1}{2}\rho_W A V^2,$$

which integrates to $V = V_0 e^{-ks}$, where m = mass of the projectile,[6] V = projectile velocity, V_0 = impact

[5] For velocities greater than 20 ft./sec., errors caused by neglect of gravity are less than 1 percent when steel spheres of $\frac{1}{4}$-in. diameter are considered.

[6] This should also include the virtual mass of the flow, but for lack of more definite information concerning the virtual mass it has not been considered in the computation. The drag coefficients obtained here are therefore slightly on the low side.

velocity, ρ_W = water density, A = maximum cross-sectional area of the projectile, $k = C_D \frac{1}{2}\rho_W A/m$, and s = depth of penetration into the water. Thus the drag coefficient can be obtained from a semilog plot of V against s. That is,

$$C_D = [m \log(V/V_0)]/(\tfrac{1}{2}\rho_W A s). \qquad (1)$$

Since the motion pictures were taken at speeds of more than 2500 frames/sec., and frame-by-frame analysis of the motion was not justified by the resolution of position on each frame, it was found most accurate to determine velocities by means of a least-square fit taken over five alternate frames. A sample semilog plot of velocity against depth is shown in Fig. 9. The drag coefficient is determined by Eq. (1) from the slope of the best fitting straight line through the calculated least-square velocity points.

Drag coefficients obtained in the preceding manner from seventeen projections of 1/4-in. spheres into water with atmosphere pressures ranging between 3 atmospheres and 1/50 atmosphere all fall within a small region about the average value, 0.27, see Table II. The standard deviation was 0.015. Hence, despite the tremendous variation in the cavity with the external pressure, the average drag coefficient is essentially constant. However a point-by-point determination of the drag coefficient might show a greater variation than was obtained by the averaging process used here.

The two-dimensional theory of the finite cavity outlined in the last section shows significant variation of drag coefficient with cavitation number. This would seem to imply that in the studied portion of the underwater trajectory of the sphere the cavitation number was effectively constant and small.

The effect of the jet striking the projectile was sometimes visible from the behavior of the $\log V$ *vs.* s plot. In these cases the straight line accurately fits the measured points until a certain depth at which the curve bends suddenly; as Fig. 9 shows, this bend was correlated with the impact of the jet on the projectile.

THEORY

The theory of two-dimensional, steady-state cavity flows has been successfully studied by means of the now classical conformal mapping methods of Kirchhoff and Helmholtz.[7] However, these flows, in which the cavity pressure equals the static pressure of the parallel flow, i.e., cavitation number equals zero, give rise to infinite cavities. A corresponding theory of the finite cavity with non-zero cavitation number is outlined here.[8]

Consider a plane parallel flow with free boundary past an obstacle, say a flat plate. Let the pressure in the cavity bounded by the free streamlines be p, the static pressure of the parallel flow at infinity P, and the undisturbed velocity of the flow V. Then from Bernoulli's equation, the condition along the free boundary is:

$$P + \tfrac{1}{2}\rho_W V^2 = p + \tfrac{1}{2}\rho_W U^2,$$

where U is the constant flow velocity along the free, bounding streamlines. By definition, cavitation number $= K =$

$$(P-p)/(\tfrac{1}{2}\rho_W V^2) = (U^2 - V^2)/V^2.$$

If $P > p$, then obviously $K > 0$, and $U > V$. When $K = 0$ the theory of Kirchhoff and Helmholtz

TABLE II. Values of drag coefficient obtained under various entry conditions.

Press (atm)	Entrance velocity	C_D
3	25	.278
2	32	.286
1 1/2	35	.288
1	39	.293
1/2	49	.278
1/2	36	.267
1/4	35	.266
1/4	36	.263
1/8	37	.248
1/8	38	.263
1/16	36	.288
1/16	37	.252
1/32	37	.295
1/32	36	.258
1/34	36	.258
1/40	47	.266
1/50	49	.286
	Average C_D	.272

[7] L. M. Milne Thompson, *Theoretical Hydrodynamics* (MacMillan Company Ltd., London, 1938).

[8] This model of the finite cavity was first suggested by H. Wagner, University of Göttingen, but was not published by him. The authors are indebted to G. Kreisel of the Admiralty Research Laboratory for acquainting them with its essential features.

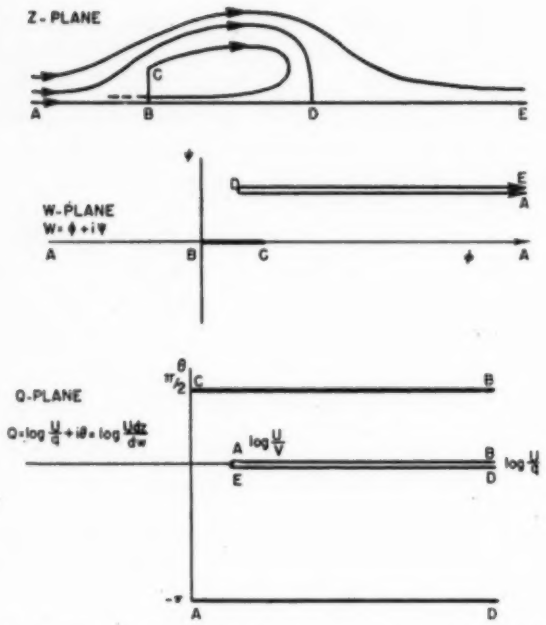


FIG. 10. Two-dimensional flow from finite cavity behind flat plate, showing flow lines in jet. The flow mapped in the W and Q planes is also shown.

applies, and gives an infinite cavity. When $K>0$, it can be shown that the conditions along the free boundary can be satisfied with a flow of the type shown in the top of Fig. 10. That is, the cavity is finite in extent, but the bounding streamlines reverse direction to form a jet infinite in length and extending "through the obstacle." Mathematically, the jet is a second sheet of a Riemann surface, and does not strike the obstacle which, of course, is not the physical situation. The method of calculating such a flow by conformal mapping techniques is outlined in the following.

The flow plane, or z-plane, is shown in Fig. 10. If $w=\phi+i\psi$ is the complex velocity potential, where $\phi(x, y)$ = velocity potential, $\psi(x, y)$ = stream function, then the upper half of the z-plane is mapped on the w-plane in the manner shown in Fig. 10; corresponding points in the two planes have the same lettering. If the actual mapping function $w=w(z)$ is known, then the entire flow is known. As is customary in flow problems with free boundaries, the distorted hodograph plane, defined by $Q=\log(U/q)+i\theta$, is introduced where $dw/dz=qe^{-i\theta}$, q = stream velocity, θ = flow direction, and hence $Q=\log(Udz/dw)$. As is readily seen, the lettered contours in the z- and w-planes are mapped onto the contour shown in the Q plane in Fig. 10. The contours in the Q- and w-planes being polygonal, the two planes can each be mapped on the same half-plane by means of the well-known Schwarz-Christoffel transformation. This gives a parametric representation of both dz/dw, and w, from which the desired features of the flow are determined by quadratures. The mathematical details are omitted here.

The drag on the obstacle can be calculated by integrating the pressures over the portion facing the stream. This tacitly neglects the effect of the jet on the drag, which is a physically reasonable assumption, since the jet is greatly weakened because of turbulent mixing at the rear of the cavity, and usually does not reach the obstacle. The results of such calculations applied to the flow past a flat plate show the drag coefficient increasing almost linearly with cavitation number so that the drag coefficient is given by $C_D=0.88(1+K)$. The value $2\pi/(4+\pi)\approx 0.88$ at zero cavitation number is the classical result for the drag coefficient of a flat plate with an infinite wake, or cavity, at its rear. These results have not yet been checked experimentally for the flat plate, but supporting evidence is available from integrated pressure measurements taken over a cavitating flat-nosed cylinder, which show a similar linear increase in drag coefficient with cavitation number.

Over the region of calculation $0<K<1.3$ the thickness of the jet at infinity was found to be almost constant—equal approximately to 0.22 the plate width.

The preceding theory sheds some light on the non-steady flow with closed cavity as observed in water-entry experiments. For a given projectile velocity and cavity pressure a unique cavity is theoretically determined. As the velocity decreases, the cavitation number rises, thereby dictating a new, smaller cavity which is inconsistent with the cavity pressure. The result is an unstable cavity which is continually shedding bubbles in an attempt to reconcile the hydrodynamic conditions with the pressure in the cavity.

DESCRIPTION OF APPARATUS

Two specially constructed, reduced-pressure tanks were used in these tests. For 1/2-in. and 1-in. models a cadmium-plated steel tank ($3\frac{1}{2}\times 3\frac{1}{2}$

×5 ft. deep) with glass sides was employed, while for 1/4-in. models a small brass tank (12×12×18 in. deep) was used. A 28-cu. ft./min. vacuum pump evacuated the tanks with negligible delay. Standard atmospheric pressure was taken as 30 in. of mercury, allowance being made in the tests for daily variations in barometric pressure. During the process of evacuation, changes in the temperature of the air within the tanks were observed by means of thermometers. The maximum temperature drop observed by this method was 5°F, rising back to within 3°F from normal in 3 or 4 minutes. Care was taken to delay all model projections for this time so that the air density within the tank could be taken as proportional to air pressure.

In order to prevent air blast and leakage from the model guns, specially designed guns were constructed for the tests. For $\frac{1}{4}$-in. and $\frac{1}{2}$-in. models, spring guns were used and found to be adequate. For 1-in. models an air gun was used, the model being driven down the barrel by air pressure acting on a light piston. At the end of the barrel the piston was stopped by a coil-spring shock absorber of annealed iron.

Pictures were taken by silhouette lighting. It was found that this method of illumination gave sharp definition of the cavity and of phenomena within the cavity. Eastman and Fastax 16-mm, high speed rotating prism-type cameras were used with quite satisfactory results. Time references in the pictures were obtained by including a synchronous motor-driven clock which could be read to 1/10 millisecond. Distance references were recorded by stadia rods mounted vertically in both tanks; the one in the large tank was divided into inches and the one in the small tank into half-inch units. Data was obtained from the motion picture films by measuring the projected images frame by frame. The order of magnitude of errors in determination of the times of surface closure was ±10 percent because of the thinness and transparency of the splashes. Times of deep closure were readily determined to within 1/2 millisecond. Errors in determination of the depths of deep closure and other distance measurements were of the order of ±0.1 in., the accuracy being limited by the grain of the film. Maximum jet velocities could not be determined any closer than about 25 percent because of the extreme thinness of the jets and because the jets were somewhat obscured by irregularities of the cavity walls.

The authors gratefully acknowledge the assistance given by Miss Jean Woodhull, and Messrs. G. F. Hickey and N. K. Stenberg in collecting and analyzing the data. We also wish to express our appreciation to Dr. J. H. McMillen for his aid in preparing this report for publication.

Statistical Aspects of Fracture Problems

BENJAMIN EPSTEIN
Coal Research Laboratory, Carnegie Institute of Technology, Pittsburgh, Pennsylvania
(Received June 16, 1947)

In recent years there has been an increasing interest in the development of statistical theories of strength. A main aim of these theories is to explain in a reasonable way such things as the dependence of the strength of specimens on their volume or length. In this paper it is pointed out that the problems posed by these models are equivalent to an important problem in mathematical statistics, namely, the distribution of the smallest value in samples of size n drawn from a population having some probability density function $f(x)$. The calculations made by mathematical statisticians give a far more complete description of the results to be expected than do the estimates to be found up to now in the technical literature.

1. INTRODUCTION

A NUMBER of articles have appeared in the literature attempting to explain on the basis of a statistical model some of the phenomena observed in connection with the fracturing of metals, textiles, and other materials under applied forces. Among other things the object of the theories is to explain in a quantitative way why one observes volume effects in the testing of materials.

In connection with some researches being carried out by the author of this paper, we have had occasion to make a partial survey of the literature on fracture. It soon became apparent that a great deal of duplication is going on and that workers in this or that special field are deriving over and over again essentially the same results. It therefore seemed worth while to write a general article which would put the subject in its proper perspective and bring together in one place some of the references scattered throughout the literature.

Another aspect of the problem was that despite the fact that the proposed theories are essentially of a purely statistical nature, most of the authors make no reference to the pertinent statistical literature. As a matter of fact, there is a vast amount of statistical theory developed which is directly applicable to this problem in the sense that it answers in an elegant way the very questions which are only crudely approximated in the technical literature.

2. BASIC FEATURES OF THE ASSUMED MODEL. THE WEAKEST LINK CONCEPT

In essence the statistical models proposed in the study of fracture take as a starting point Griffith's theory[1] which states that the reason for the difference between the calculated strengths of materials and the actual observed values resides in the fact that there exist flaws in the body which will weaken it. This means then that there will be a distribution of strengths in a given specimen in the sense that a different amount of force will be needed to fracture a specimen at one or another point. Of course, this is purely conceptual in nature since one cannot actually test the strength of an isolated element without changing the conditions which exist when the element is actually in the body. If one assumes that the flaws are distributed at random with a certain density per unit volume then the statistical formulation of the problem becomes apparent. If the flaw concept is accepted then the strength of a given specimen is determined by the weakest point in the specimen or by the smallest value to be found in a sample of size n where n is the number of flaws. Clearly n increases as the volume increases and, therefore, the problem of finding out how the strength depends on the volume of the sample is equivalent statistically to studying the distribution of the smallest value as a function of n, the sample size. This statistical problem is an important one on which much theoretical work has been done. More will be said about this later in the paper.

As far as is known to the present author, the

first worker to realize the close connection between the strength of a specimen and the distribution of smallest values was F. T. Peirce[2] of the British Cotton Industry Research Association. To use his own words: "It is a truism, of which the mathematical implications are of no little interest, that the strength of a chain is that of its weakest link." This sentence states in compact form the essence of the assumed model.

Peirce was led to this problem in connection with some researches on the construction of reliable procedures for the testing of yarn. In order to accomplish this primary aim it was essential to study the effect of certain variables which one might expect on *a priori* grounds to be capable of greatly influencing the results unless properly controlled. Among the factors considered was the length of the specimen. It was observed, for example, that if specimens of some fixed length l broke under normally distributed loads, then specimens of length nl, where n is an integer greater than one, will break under loads whose distribution is negatively skewed the larger the value of n. Peirce wished to account for this phenomenon, which is incidentally a characteristic of distributions of breaking strength and thus was led to the formulation of his "chain" model and to the equivalent statistical problem. It is interesting that the paper by Peirce was published at the time when the first basic statistical researches on the distribution of extreme values were carried out by the British statistician, Tippett.[3] Peirce mentions explicitly in his paper that the theoretical work of Tippett is of importance in obtaining quantitative results concerning the distribution of breaking strengths.

The application of essentially the same ideas to the study of the strength of materials is found in two papers by the Swedish engineer, Weibull.[4] Unlike Peirce who assumed a Gaussian distribution of strengths, he makes the assumption that $F_0(\sigma)$,* the probability of breakage of a unit volume as a function of the stress σ is given by $F_0(\sigma)=1-\exp[-(\sigma/\sigma_0)^m]$ where σ_0 and m are unknown parameters which may depend on the characteristics of the material under test. As mentioned above, the term probability of breakage of a unit volume is purely conceptual. Again he is interested in finding out how the probability of rupturing a specimen of volume V depends on V. Weibull shows that the strength of specimens of volume V is proportional to $\sigma_0 V^{-(1/m)}$. It is apparent from his calculations that one of the primary motivations for his particular choice of $F_0(\sigma)$ is to make it easier to obtain results in closed form. Corroboration to some extent of Weibull's theory has recently been demonstrated by Davidenkow, Shevandin, and Wittman.[5]

Shortly after Weibull's work appeared the Russian physicists, Kontorova[6] and Frenkel and Kontorova[7] published papers which again reiterate the "weakest link" concept. They consider crystalline specimens with flaws distributed at random throughout the specimens, thus giving a distribution of strengths throughout the specimen. They assume, as did Peirce in the case of testing the strength of yarn, that these strengths are normally distributed. The main result of the Frenkel-Kontorova paper is that if the distribution of strengths x is given by the probability density function**

$$f(x)=\frac{1}{(2\pi)^{\frac{1}{2}}\sigma}\exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right],$$

then the strength of specimens of volume V large enough to contain many flaws is approximately given by $\mu-(2\sigma)^{\frac{1}{2}}(\log nV-\log 2(\pi)^{\frac{1}{2}})^{\frac{1}{2}}$, where n is the average number of flaws per cubic centimeter of the material. Frenkel and Kontorova leave the impression that they consider this result to be very significant and to justify the assumption that the strengths throughout the body due to flaws have a Gaussian distribution. We shall see later in the statistical part of this paper how their result and many others are contained as special cases of a general and well-known statistical theory of the distribution of extreme values.

It is interesting that Frenkel and Kontorova obtained a good approximation to the most probable value (or mode) of the smallest value in samples of size n (or volume V) drawn from a normal distribution without realizing that a more

* $F_0(\sigma)$ is in statistical terminology a cumulative distribution function which is monotone non-decreasing with $F_0(0)=0$ and $F_0(\infty)=1$. $F_0(\sigma)$ is the fraction of unit volumes breaking under stresses $\leqslant\sigma$. The probability density function is: $\frac{dF_0(\sigma)}{d\sigma}$.

** σ is the standard deviation of the normal distribution $f(x)$.

general theory existed which would not only give them the most probable value, but would also give the distribution of the smallest values to be found in samples of size n. What is more, this theory permits one to state in a quantitative way what will happen for distributions which are not normal. If Frenkel and Kontorova had seen the problem in its general statistical context they would not have made some of the statements contained in the last two pages of their paper. As an example of what we mean, it was seen earlier that they obtained the result that there is a linear relationship between the most probable value of the strength of specimens of volume V and $(\log V)^{\frac{1}{2}}$ if the existence of flaws is assumed to lead to a Gaussian distribution of strengths throughout the specimen. But it can be shown that if the strengths x due to flaws have a Laplace distribution given by $1/2\lambda \exp[-|x-\mu|/\lambda]$, where μ and λ are parameters playing roles similar to the μ and σ in the Gaussian case, then the most probable value of the strength in specimens of volume V is linearly related to $\log V$ by the relationship $\mu-\lambda \log(nV/2)$, where n is the average number of flaws per cubic centimeter of the material. It does not seem as if the present state of knowledge about the dependence of strength on volume is sufficiently precise to be able to permit one to choose between a dependence on $\log V$ or $(\log V)^{\frac{1}{2}}$, if attention is confined only to the most probable value of strength in specimens of volume V. We shall see, in our statistical discussion, that there is an important difference between the *distributions* of smallest values obtained from the Gauss distribution on the one hand and the Laplace distribution on the other, and that based on such considerations there actually may exist grounds for preferring one distribution function over another. However, all these considerations are purely academic as long as no experimental evidence exists as to the number, distribution or severity of flaws. Possibly some method of counting and classifying flaws may be developed which will help in deciding which *a priori* distributions are to be assumed.

Frenkel and Kontorova went to considerable length to emphasize that their approach is to be preferred to that of Weibull because his assumption for $F_0(\sigma)$ is devoid of physical meaning and because he does not realize that he is dealing with a situation where essentially it is the weakest link which determines the breaking strength of a specimen. If one views the problem in its general statistical aspects, the criticisms of Weibull's work are hardly justified. In the first place, present knowledge is still not sufficiently precise to say whether or not $F_0(\sigma)$ is less suitable as an *a priori* strength distribution than one which is purely Gaussian. Both assumptions involve two free undetermined parameters and probably give enough freedom to fit existing data reasonably well on either hypothesis. In the second place, Weibull does implicitly involve the principle that it is the weakest link which determines fracture strength. It is easy to calculate the maximum (mode or most probable value) of the distribution of strengths of samples of size n (or specimens of volume V) drawn from each of the distributions in question, and one can verify that results of both the Russian and Swedish investigators are good approximations of the respective maxima appropriate to their *a priori* distribution assumptions.

All the papers discussed up to this point were based essentially on the concept that fracture is determined by the weakest link. It is interesting to mention a recent paper by H. E. Daniels[8] which deals with the statistical theory of the strength of bundles of threads. This problem is much more complicated and differs in the important respect that in the one case one is dealing conceptually with the breakage of a long chain linked in series and, where the force on each link is equal to the force applied to the chain as a whole. In Daniels' problem the load is applied to a set of elements in parallel so that the application of a load S to a bundle containing n threads will throw a load equal to S/n upon each thread. It turns out that a bundle will not break under load S if, and only if, there exists an integer $k \leqslant n$ such that k among the threads have strengths exceeding S/k. The problem of the distribution of strength of bundles of n threads can no longer be treated as a problem in the distribution of smallest values. It turns out that, whereas the distribution of strengths of single long strands is negatively skewed, the distribution of strengths of bundles of n threads is asymptotically normal for large n under fairly general assumptions

about the distribution of strengths of individual strands.

We mention Daniels' result, even though it is not a purely "weakest link" problem, because of its possible importance in studying certain aspects of the strength of materials. In any real problem the situation is probably neither that of elements in series or in parallel, but rather elements distributed in some rather complicated arrangement. It is, however, valuable to know what to expect in the case where certain limiting idealized models are assumed to be valid.

3. THE STATISTICAL THEORY OF THE DISTRIBUTION OF EXTREME VALUES

We have mentioned several times that the "weakest link" problem is precisely equivalent to that of the distribution of smallest values in samples of size n. In this section a survey of the literature in this field will be given and it will be pointed out how the results may be used in studying the fracture problem.

During the past twenty years a considerable literature has been developed directly concerned with the problem of studying the distribution of extreme values (i.e., smallest or largest values) in samples of size n drawn from a population possessing probability density function $f(x)$. More generally, the literature treats the distribution of the νth values from the top or bottom, when one arranges the n sample values in the order of their magnitude. The exact distribution of extreme values for both small and large values of n, under the assumption that the underlying population is Gaussian, has been studied by a number of writers and some of the results have been reduced to tables.[9] There are almost no results for small values of n if $f(x)$ is not normal.

Insofar as applications of the statistical theory to the fracture problem are concerned, one is interested primarily in the distribution of the smallest value in samples of size n for large values of n, and it is fortunate that an extensive asymptotic theory is available which gives in implicit form the distribution of the largest, smallest, or νth values in large samples ($n \to \infty$) for any $f(x)$, subject to trivial analytic restrictions. There are a number of interesting cases for which it is possible to find explicitly the distributions of extreme values.

The basic work on the asymptotic theory was done by R. A. Fisher and Tippett[10] and Gumbel.[11] The latter author has also made some very interesting and important applications of the distribution of extreme values to the study of floods,[12] the distribution of ages in mortality tables,[13] and to other physical phenomena. An excellent account of the basic theory as it exists at the present time will be found in a recent book by H. Cramér.[14]

The statistical theory underlying the distribution of smallest values is contained in six fundamental formulae (Eqs. (1) to (6) below). These formulae permit one to answer a number of questions which are of direct physical interest in connection with the fracture problem. Suppose that the underlying distribution is given by the continuous probability density function $f(x)$ with associated cumulative distribution function

$$F(x) = \int_{-\infty}^{x} f(x)dx,$$

then the distribution of the smallest value in samples of size n drawn from the population is given by the probability density function:

$$g_n(x) = nf(x)(1 - F(x))^{n-1}, \tag{1}$$

or by the cumulative distribution function:

$$G_n(x) = \int_{-\infty}^{x} g_n(x)dx = 1 - (1 - F(x))^n. \tag{2}$$

Equations (1) and (2) already permit one to give graphical descriptions of $g_n(x)$ and $G_n(x)$ for given $f(x)$, and it is easy to find the mode of $g_n(x)$ or the *most probable value* of the smallest value in samples of size n by finding the maximum of $g_n(x)$, i.e., solving the equation $g_n'(x) = 0$. If a solution exists, it may be written in the form

$$f^2(x_n^*)(n-1) = [f'(x_n^*)][1 - F(x_n^*)]. \tag{3}$$

If the initial distribution is limited, the distribution of the smallest value decreases monotonically, and no mode in the proper sense exists. Formula (3) is the one used either implicitly or explicitly by all of the writers treating the fracture process as a "weakest link" phenomenon and the results of each of these writers differ simply because x_n^*, the most probable value of

TABLE I.

Probability density function	x_n^*, Mode of smallest value in samples of size n
1. Rectangular distribution $f(x)=1/b-a, \quad a \leqslant x \leqslant b$ $f(x)=0, \quad$ elsewhere	a
2. Cauchy distribution $f(x)=\frac{1}{\pi}\frac{\lambda}{\lambda^2+(x-\mu)^2}$	$\mu-\frac{\lambda n}{2\pi}$
3. Laplace's distribution $f(x)=\frac{1}{2\lambda}\exp-\frac{\lvert x-\mu\rvert}{\lambda}$	$\mu-\lambda\log\frac{n}{2}$
4. Gauss' distribution $f(x)=\frac{1}{(2\pi)^{\frac{1}{2}}\sigma}\exp-\frac{(x-\mu)^2}{2\sigma^2}$	$\mu-\sigma(2\log n)^{\frac{1}{2}}$ $+\sigma\frac{\log\log n+\log 4\pi}{2(2\log n)^{\frac{1}{2}}}$
5. Distribution of type assumed by Weibull $f(x)=\alpha\beta x^{\beta-1}\exp(-\alpha x^{\beta}), \quad x\geqslant 0,$ $\alpha>0, \beta>1$ $F(x)=1-\exp(-\alpha x^{\beta})$	$\frac{1}{[\alpha n]^{1/\beta}}(1-1/\beta)^{1/\beta}$

If we introduce the mode x^* of the initial Weibull distribution, the mode of the smallest value becomes $x^* n^{-1/\beta}$.

the smallest value in samples of size n, depends on the form of $f(x)$, the underlying probability density function from which the samples are drawn. The way in which x_n^* depends on $f(x)$ is perhaps brought out best by a table in which $f(x)$ is assumed to be one of a number of common distribution functions. The results are given in Table I.

TABLE II.

Probability density function	Distribution of smallest values in samples of size n (for n large)
1. Rectangular distribution in interval $(a \leqslant x \leqslant b)$	$x=a+(b-a)\xi/n$
2. Cauchy disrtibution with parameters $(\mu;\lambda)$	$x=\mu-\frac{\lambda n}{\pi\xi}$
3. Laplace distribution with parameters $(\mu;\lambda)$	$x=\mu-\lambda\log\frac{n}{2}+\lambda\log\xi$
4. Gaussian distribution with parameters $(\mu;\sigma)$	$x=\mu-\sigma(2\log n)^{\frac{1}{2}}$ $+\sigma\frac{\log\log n+\log 4\pi}{2(2\log n)^{\frac{1}{2}}}$ $+\frac{\sigma}{(2\log n)^{\frac{1}{2}}}\log\xi$
5. Weibull distribution with parameters $(\alpha;\beta)$	$x=(\xi/n\alpha)^{1/\beta}$

ξ is distributed with probability density function $h(\xi)=e^{-\xi}, \quad \xi\geqslant 0.$

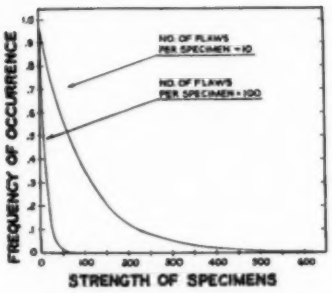


FIG. 1. Frequency distribution of strengths of specimens containing various number of flaws and for which the flaw strengths are rectangularly distributed with $a=0$, $b=1000$.

Equations (1), (2), and (3) are not enough if one wishes to find not only the most probable value of the smallest value in samples of size n, but also the actual distribution of values. In none of the papers dealing with the fracture problem on a statistical basis does one find a consideration of this important question. It is precisely at this point that the work of mathematical statisticians gives the answers desired. In order to study the distribution of smallest values it is convenient to introduce a new vari-

TABLE III.

Probability density function	Remarks on general characteristics of the distribution of x, the smallest value in samples of size n (n large)
1. Rectangular distribution in interval (a, b)	x is nearly independent of n. $D^2(x)=(b-a)^2/n^2$
2. Cauchy distribution with parameters $(\mu;\lambda)$	Most probable value of x decreases linearly with n. Both mean value and variance are infinite
3. Laplace distribution with parameters $(\mu;\lambda)$	Most probable value decreases as a multiple of $\log n$. Variance remains equal to $\lambda^2\pi^2/6$ independently of n
4. Gaussian distribution with parameters $(\mu;\sigma)$	Most probable value decreases as a multiple of $(\log n)^{\frac{1}{2}}$. Variance decreases as n increases and is given by $\pi^2\sigma^2/12\log n$
5. Weibull distribution with parameters $(\alpha;\beta)$	Most probable value decreases as $n^{-1/\beta}$. Mean value decreases as $n^{-1/\beta}$ and variance as $n^{-2/\beta}$

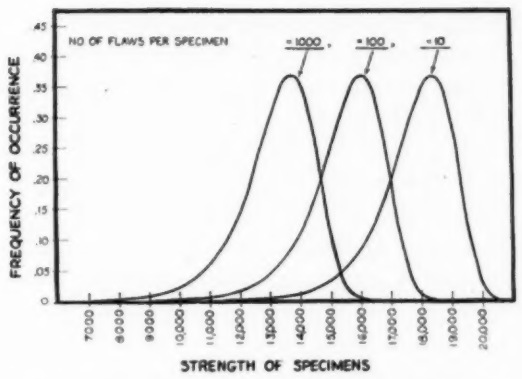


FIG. 2. Frequency distribution of strengths of specimens containing various numbers of flaws and for which the flaw strengths are distributed according to the Laplace distribution with $\mu = 20{,}000$, $\lambda = 1000$.

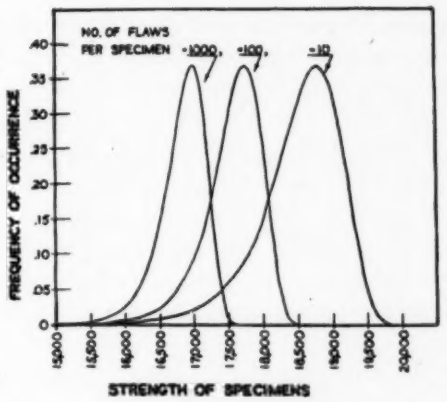


FIG. 3. Frequency distribution of strengths of specimens containing various numbers of flaws and for which the flaw strengths are normally distributed with $\mu = 20{,}000$, $\sigma = 1000$.

able defined as:

$$\xi = nF(x). \tag{4}$$

In terms of this variable, (1) becomes:

$$h_n(\xi) = \left(1 - \frac{\xi}{n}\right)^{n-1}. \tag{5}$$

$$h(\xi) = \lim_{n\to\infty} h_n(\xi) = e^{-\xi} \tag{6}$$

is the probability density function of ξ as: $n \to \infty$.

Using Eqs. (4) and (6), one can now completely specify the distribution of x, the smallest value in samples of size n for large values of n.

The distribution of the smallest value in samples of size n for the various population probability density functions of Table I are given in Table II.

The results in Table II specify completely how x, the smallest value in samples of size n, is distributed for various common probability density functions. The probability density function of ξ, $h(\xi) = e^{-\xi}$ is so simple that it is an easy matter graphically to see how x is distributed about its most probable value. In order to form quantitative judgments about the mean values of x and the dispersion of x as measured by the variance of x, it is useful to calculate $E(x)$ and $D^2(x)$ in each case. E and D are symbols meaning expected value of x and variance of x, respectively. If x is a random variable distributed with the probability density function $f(x)$, then $E(x)$ and $D^2(x)$ are defined by the equations:

$$E(x) = \int_{-\infty}^{\infty} x f(x)dx \tag{7}$$

and

$$D^2(x) = E(x^2) - [E(x)]^2 = \int_{-\infty}^{\infty} x^2 f(x)dx - \left[\int_{-\infty}^{\infty} x f(x)dx\right]^2. \tag{8}$$

Since the random components of x in Table II are functions of ξ only, this means that one should calculate such quantities as $E(\xi)$, $E(1/\xi)$, $E(\log\xi)$, and $E(\xi^{1/\beta})$, and also the corresponding D^2. If this is done, then the results will be as shown in Table III.

It is clear how one can interpret Table III physically if one is interested in finding out the consequences of various assumptions about the distributions of strengths due to flaws. For instance, a rectangular assumption would imply that no dependence on volume exists; a Cauchy distribution would lead to results which are physically unreasonable. Assumptions (3), (4), and (5) all imply that the strength does decrease with increasing volume. Assumptions (3) and (4) lead to relationships which are semi-logarithmic in character ($\log V$ in the Laplace case, $(\log V)^{\frac{1}{2}}$ in the Gauss case) with the important difference that in the Laplace case the distribution does not become narrower with increasing V, whereas it does in the Gauss case.

It is interesting to give a few numerical examples to illustrate the dependence of strength

on the size of specimen (or the number of flaws under various distribution assumptions). As a first example, consider a situation where the flaw strengths (measured in what ever units are appropriate to the problem) are rectangularly distributed between 0 and 1000, i.e., $a=0$, $b=1000$. In Fig. 1 is given the distribution of strength (or smallest values) in samples containing $n_1=10$ and $n_2=100$ flaws drawn from the given rectangular distribution. As a second example, consider a situation where the flaw strengths are distributed according to the Laplace distribution with $\mu=20{,}000$ and $\lambda=1000$. In Fig. 2 is given the distribution of strength in samples containing $n_1=10$, $n_2=100$, and $n_3=1000$ flaws drawn from the Laplace distribution. As a third example, let the flaw strength be normally distributed with $\mu=20{,}000$ and $\sigma=1000$. In Fig. 3 is given the distribution of strengths in samples containing $n_1=10$, $n_2=100$, and $n_3=1000$ flaws drawn from the normal distribution. All of these graphs may be readily constructed from the distribution formulae given in Table II. It may be of interest to the reader to construct similar graphs for the Cauchy, Weibull, and various other distributions. It should be mentioned that the graphs in Figs. 1 and 3 have not been normalized so as to give the same area under each distribution curve.

4. CONCLUDING REMARKS

In this section we wish to sum up, amplify, and point out some of the implications of the contents of this paper. When one looks at the problem of fracture in its most general setting it is clear that the "weakest link" concept, when applicable, may be the key to the occurrence of certain observed relationships between the strength of specimens and their size. The term strength may in some cases mean mechanical strength, or electrical strength, or ability to stop the passage of light rays, or may perhaps be the life span (or interval between repairs) of a device which ceases to function when any one of a number of vital parts breaks down. From a statistical point of view, certain phenomena, which may on the surface appear to be different, are really equivalent since they lead in each instance to the same problem, namely, the distribution of the smallest value in large samples. Now, it is not unreasonable to suppose that the distribution of strengths due to flaws is Gaussian or perhaps more generally of the form $A\exp(-B|x-\mu|^p)$ for large values of $|x-\mu|$ (where A, B, μ, and p are positive constants). Under such hypotheses statistical theory leads to the prediction that the most probable value of the smallest value in samples of size n must decrease as $(\log n)^{1/p}$ and that the distribution of smallest values in samples of size n must be negatively skewed. This means that given a physical phenomenon for which the weakest link hypothesis is a reasonable approximation to the true state of affairs, then under certain distribution assumptions about the strengths of the links there is a *statistical compulsion* for a skewed distribution of strengths of samples of the same size and an approximately semi-logarithmic dependence of the most probable value of the strength (or whatever is equivalent to strength) on the size of the specimen or number of places at which a break may occur. This may account for the fact that workers in many different fields have observed skewed distributions and approximately semi-logarithmic relationships in studies on the strength of materials.

Many scientists accept the fact that the Gaussian distribution plays a fundamental role in science and, in fact, there are many who feel that this is the only distribution which nature calls truly her own. We have shown in this paper, however, that in a certain class of phenomena the characteristic distributions are far from normal and are, in fact, strongly skewed to the left. There are still other instances which lie outside the scope and purpose of this paper where still other "characteristic" distributions are found. This is particularly true of the particle size distributions arising from crushing and grinding where one finds a marked tendency for the logarithms of the particle size to be normally distributed.

One other important point that we wish to emphasize is that one should not expect the "weakest link" hypothesis, attractive as it may be, to explain all phenomena relating to the strength of materials. In the first place it is implicitly assumed in this paper that the flaws are all independent and do not in any way influence each other. There are physical situations where such an assumption is not justified.

For example, if the surface of a beam subjected to bending is notched, the strength may be diminished many fold. However, if many notches are put on the surface, the strength is hardly affected. This finds an important application in the fatigue strength of a body having a highly polished surface. A single scratch may be ruinous whereas roughing up the whole surface with many scratches may bring about little change in the strength.

In the second place there have been attempts (see, for example, A. M. Freudenthal[15]) to apply essentially the weakest link concept to the study of the fatigue of materials. We believe that such an attack is basically incorrect even if it leads to the prediction of relationships which agree reasonably well with what is found in nature. For example, the relationship between the length of time required to break a specimen and the applied load is in its very essence a time-dependent problem. The specimen is changing in time and its distribution of strengths due to flaws is also changing and, therefore, any essentially static approach which uses the "weakest link" concept without modification leaves out certain basic features of the process. It does not seem correct, in the treatment of fatigue problems, to put time on the same level with length or volume, and to argue that n seconds is to one second as n centimeters is to one centimeter in the sense that if $1-p$ is the probability of an event not occurring in a specimen one centimeter long (or in a one second time interval), then $(1-p)^n$ is the probability of the event not occurring in a specimen n centimeters long (or in an n second time interval). Certain aspects of the question of fatigue are to be studied in another paper.

5. ACKNOWLEDGMENT

It is a pleasure to thank Professor Edward Saibel of the Department of Mathematics of the Carnegie Institute of Technology for his interest and helpful suggestions in connection with this paper.

BIBLIOGRAPHY

1. A. A. Griffith, "The phenomena of rupture and flow in solids," Phil. Trans. Roy. Soc. **221A**, 163 (1920); "The theory of rupture," Proc. Int. Cong. App. Mech., Delft **1**, 55 (1924).
2. F. T. Peirce, "Tensile tests for cotton yarns *v.* 'The weakest link'—theorems on the strength of long and of composite specimens," J. Tex. Inst. **17**, 355 (1926).
3. L. H. C. Tippett, "On the extreme individuals and the range of samples taken from a normal population," Biometrika **17**, 364 (1925).
4. W. Weibull, "A statistical theory of the strength of materials," Ing. Vetenskaps Akad. Handl. No. 151 (1939); "The phenomenon of rupture in solids," *Ibid.* No. 153 (1939).
5. N. Davidenkow, E. Shevandin, and F. Wittman, "The influence of size on the brittle strength of steel," J. App. Mech. **14**, A63 (1947).
6. T. A. Kontorova, J. Tech. Phys. U.S.S.R. **10**, 886 (1940).
7. J. I. Frenkel and T. A. Kontorova, "A statistical theory of the brittle strength of real crystals," J. Phys. U.S.S.R. **7**, 108 (1943).
8. H. E. Daniels, "The statistical theory of the strength of bundles of threads I," Proc. Roy. Soc. (London) **183**, 405 (1945).
9. K. Pearson, *Tables for statisticians and Biometricians* (Cambridge University Press, Teddington, England, 1924 and 1931), Vol. I, second edition and Vol. II.
10. R. A. Fisher and L. H. C. Tippett, "Limiting forms of the frequency distribution of the largest and smallest member of a sample," Proc. Camb. Phil. Soc. **24**, 180 (1928).
11. E. J. Gumbel, "Les valeurs extrêmes des distributions statistiques," Ann. inst. Henri Poincaré **5**, 115 (1936).
12. E. J. Gumbel, "The return period of flood flows," Ann. Math. Statistics **12**, 163 (1941).
13. E. J. Gumbel, "La durée extrême de la vie humaine," *Actualités Scientifiques et Industrielles* (Hermann et Cie, Paris, 1937).
14. H. Cramér, *Mathematical methods of statistics* (Princeton University Press, Princeton, New Jersey, 1946).
15. A. M. Freudenthal, "The statistical aspect of fatigue of materials," Proc. Roy. Soc. (London) **A187**, 416 (1946).

An Electronic Differential Analyzer

J. S. KOEHLER
Physics Department, Carnegie Institute of Technology, Pittsburgh, Pennsylvania
(Received July 14, 1947)

An electrical device is designed which will solve ordinary, non-linear, non-homogeneous differential equations or the boundary value problems based on such differential equations. The device is based on the fact that the charge on the condenser in a series resonant circuit varies with the time in a manner described by a linear second-order differential equation. The desired variations of the coefficients with time are fed into the apparatus by using a variable voltage generator. This generator provides a small voltage which varies with time in accordance with any given curve. The variable voltage generator has been constructed and operates satisfactorily. The resulting solution is drawn by an oscillograph. The instrument is designed to furnish a solution which does not deviate from the true solution by more than four percent.

INTRODUCTION

THAT there exists a very great need for devices which will solve differential equations and boundary value problems is evident from the time and energy spent in the development of the ENIAC. These large instruments are accurate and powerful, but there are a large number of problems which would be adequately treated by a simpler less accurate instrument. It should also be noted that if a solution for a given problem is known with, say, one percent accuracy, then there exist numerous techniques which enable one to obtain a solution having an accuracy of one-tenth of a percent without too much effort. It is the initial rough solution which is often very difficult to obtain.

The present paper describes the design of an electronic analog of an ordinary differential equation. The instrument can also be used to solve one-dimensional boundary value problems. The design given enables an answer to be obtained with an accuracy of plus or minus four percent.

The essential idea used to obtain the analog is that the behavior of the dependent variable of a second-order ordinary differential equation with variable coefficients can be imitated by the behavior of the charge on the condenser of a series resonant circuit. It is, of course, necessary that the resistance, the capacitance, and the inductance of the circuit be made to vary in a predetermined manner with time. The principal difficulty in obtaining a practical circuit is that the time constants of the circuit are usually too short to permit the necessary switching. This paper gives the design of a circuit having suitable time constants. Various parts of the circuit have been experimentally tested, but the complete apparatus has never been assembled.

The apparatus includes a circuit used to supply a voltage which varies with the time in accordance with a given curve. This circuit has been tested and operates in a satisfactory fashion.

GENERAL DESCRIPTION

The circuits required will first be described in a rather special case, and then the modifications required in more general cases will be discussed.

Let q be the dependent variable. The independent variable will be called t. Assume that at $t=0$ the values of q and dq/dt are given. In addition, suppose that q varies with t in accordance with the differential equation,

$$(d^2q/dt^2)+A(t)(dq/dt)+B(t)q=0, \qquad (1)$$

where A and B are known functions of t (i.e., a graph of A *versus* t and a graph of B *versus* t should be available).

Now consider the simple series resonant circuit shown in Fig. 1. In the figure, R is the resistance of the circuit, L is the self inductance of the circuit, and C is the capacitance of the circuit. The charge q on one plate of the condenser C is determined by the initial conditions and by the equation,

$$(dq/dt^2)+(R/L)\cdot(dq/dt)+(q/LC)=0, \qquad (2)$$

where t is the time. The similarity between (1) and (2) is evident. The circuit to be described

operates as follows:

1. The initial conditions are established.
2. Then the values of R and C are varied with time in such a way that $A(t)=R/L$ and $B(t)=1/LC$.
3. The value of q is traced as a function of time on a cathode ray screen.
4. The entire cycle is repeated thirty times a second. Persistance of vision enables one to see a fixed curve on the screen. This curve can be photographed if a permanent record is desired.

The actual circuit used is shown in the block diagram given in Fig. 2. P_2 and P_3 are 929 vacuum photo-tubes; C' is in all cases a 0.5 μf, 400 volt condenser; $\rho=33$ megohms, $R_1=12.5$ megohms, $L=0.0042$ henry; $C_3=50$ μf (500 volts); C_1 is variable from 0.215 μf to 0.225 μf; and C_2 is a special double condenser which will be described later. AC Amp 1, AC Amp 2, and AC Amp 3 are three different AC amplifiers. In amplifiers 1 and 2 the grid bias of one or more of the tubes is varied by applying a voltage at the $BIAS$ terminal. CR_1, CR_2, and CR_3 are cathode ray oscilloscopes. The square wave generator provides external synchronization for the horizontal sweep in all three cases. The *q versus t* trace appears on CR_1. The *CHARGE* box is an

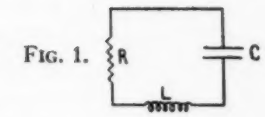


FIG. 1.

electronic switch actuated by the square wave generator. The condensers are charged periodically through this circuit, the power being supplied by battery B.

The operation of the instrument is best understood if the action of the variable condenser, the variable resistance, the charging circuit, and the *BIAS* voltage supply are described and then it is shown how these circuits are operated in coordination in the differential analyzer.

THE VARIABLE CONDENSER

The quantity $1/C$ appears in the third term of Eq. (2). The circuit used should be able to oscillate at frequencies which are much higher and much lower than the thirty cycles per second which is the repetition frequency of the instrument. If the frequency of the resonant circuit be variable from 6~/sec. to 1000~/sec., then it is found that with the value of L given above, the largest value of C necessary is 0.17 farad while the smallest value needed is 6×10^{-6} farad.

The variable condenser is constructed as shown in Fig. 3. k is the dielectric constant. The AC

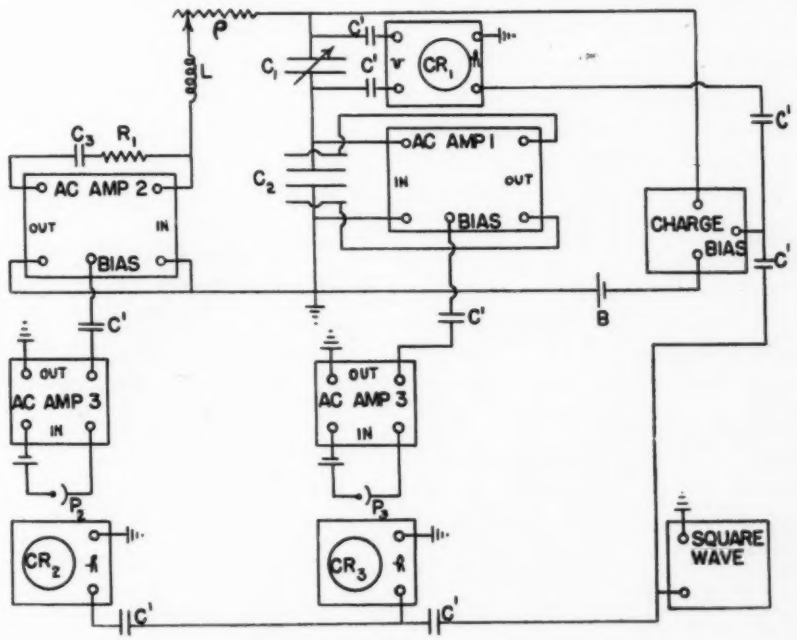


FIG. 2.

amplifier shown has an input resistance of 10^8 ohms. Suppose that the input voltage to the amplifier is V volts and the output voltage is nV volts. Suppose also that the output voltage can be varied from n_0 to $n_0+\Delta$ by changing the grid voltages in one or more of the tubes in the amplifier. If the output voltage is out of phase with the input voltage, the effective capacitance of the inner pair of plates is obtained as follows: σ_i is the surface charge density on the ith conductor (see Fig. 4). By Guass' law we have: $D'=4\pi\sigma_2$, $E'=(4\pi\sigma_2)/k_2$, $D'-D=-4\pi\sigma_1$, and $E=D/k_1$, where the k's are dielectric constants as shown in Fig. 4. If A is the area of each of the four plates and a is the distance between plates then:

$$\frac{1}{C_2}=\frac{V}{\sigma_1 A}=\frac{Ea}{\sigma_1 A}=\frac{-8\pi a}{A[(n-1)k_2-2k_1]}.$$

Remembering that $n=n_0+\Delta$ where Δ is small, the above equation is:

$$\frac{1}{C_2}=\frac{-8\pi a}{A}\left[\frac{1}{(n_0-1)k_2-2k_1}-\frac{\Delta k_2}{\{(n_0-1)k_2-2k_1\}^2}+\frac{\Delta^2 k_2^2}{\{(n_0-1)k_2-2k_1\}^3}+\cdots\right]. \quad (3)$$

If the dimensions of the condenser plates are given in centimeters, then C_2 in Eq. (3) is in electrostatic units. Note that C_2 is negative when Δ is zero, and $1/C_2$ becomes more positive as Δ increases. The double condenser should be constructed in such a way that the time taken for a charge on the inner plates to leak through the input resistor of the amplifier should be large in comparison with the period of operation of the amplifier. This requires that the inner plates taken by themselves should have as large a capacitance as possible. We have therefore taken $a=1$ mm, $A=200$ cm^2, $k_1=9$.

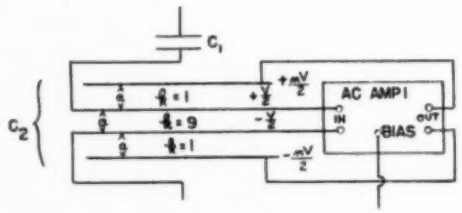


FIG. 3.

It is next necessary to arrange matters so that higher order terms in (3) can be neglected. We have taken $\Delta_{max}=100$ and $n_0=2501$. In this case the ratio of successive terms in (3) is 0.04.

The condensers C_1 and C_2 are in parallel so that if C is the capacitance of the combination, then $1/C=1/C_1+1/C_2$. Put

$$\frac{1}{C_1}=\frac{8\pi a}{A}\left[\frac{1}{(n_0-1)k_2-2k_1}-\frac{\frac{1}{2}\Delta_{max}k_2}{\{(n_0-1)k_2-2k_1\}^2}\right], \quad (4)$$

then

$$\frac{1}{C}=-\frac{8\pi a k_2[\frac{1}{2}\Delta_{max}-\Delta]}{A\{(n_0-1)k_2-2k_1\}^2}, \quad (5)$$

so that $1/C$ varies linearly from a negative value to an equal positive value as Δ varies from zero to Δ_{max}. If $k_2=1$, then the smallest value of C possible is 10.9×10^{-6} farad which yields for the maximum frequency of our resonant circuit 800~/sec. The value of C_1 given by (4) is 2.19×10^{-7} farad. The charges on C_1 and C_2 are equal and this charge is measured by CR_1.

The voltages V_1 and V_2 across C_1 and C_2 are related as follows:

$$V_1/V_2=C_2/C_1. \quad (6)$$

The maximum input voltage to the amplifier is 0.064 volt. Using this for V_2, the smallest value of V_1 possible can be calculated from Eq. (6) by inserting the minimum value of C_2. One finds that $V_1=0.065$ volt. This voltage is sufficient to deflect the beam in a Dumont 208 oscilloscope completely across the screen if maximum gain is used in the oscilloscope.

The design adopted for amplifier 1 is given in Fig. 5. Experimentally it is found that this amplifier with an input voltage of 0.064 volt has a mid-band amplification of 2300. The amplification drops by 2 percent in going from 20 to 2000~/sec. and it drops by 14 percent in going from 200 to 2000~/sec. A change of 0.08 volt in the bias voltage changes the amplification by

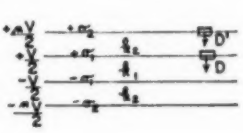



FIG. 4.

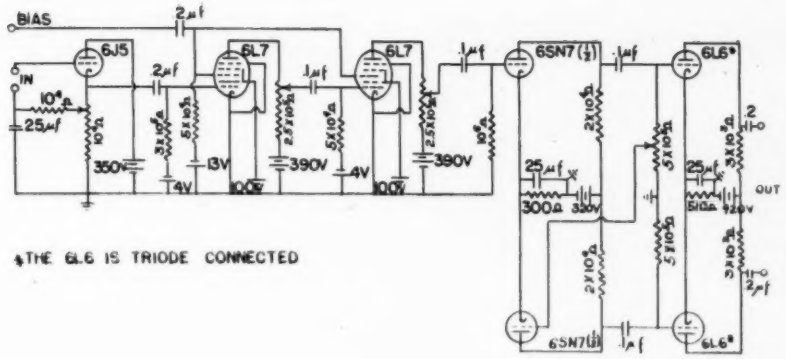


FIG. 5.

a hundred. The amplifier is linear for changes in both the input voltage and in the *BIAS* voltage over the ranges of voltage of interest for the present application.

THE VARIABLE VOLTAGE GENERATOR

If our electrical circuit is to imitate the behavior described by the differential Eq. (1), then the coefficient $1/LC$ must vary with time as $B(t)$ in Eq. (1) varies with t. This is accomplished by constructing a circuit which furnishes a voltage which varies with time in any desired fashion. If this voltage is then made to vary with time as $B(t)$ varies with t, then upon application of the voltage to the *BIAS* terminal of amplifier 1 the desired variation of the coefficient $1/LC$ is achieved.

The variable voltage generator is made up of a cathode ray oscilloscope, an opaque diaphram which has a properly shaped opening in it, and a photoelectric cell whose output is usually amplified. The circuit functions as follows: The spot on the cathode ray screen is swept vertically very rapidly (i.e., with a frequency of 200 kilocycles per second or more). A second sweep circuit moves the spot horizontally with a frequency of 30 cycles per second. If the decay time of the phosphor used on the cathode ray screen is long compared with the time required for a single vertical sweep, but short compared with the time required for a horizontal sweep then this arrangement will produce a vertical line of uniform brightness which moves across the screen in a thirtieth of a second. (The same effect could be achieved by designing a cathode ray tube which produces a uniform line spot rather than the usual point spot).

The photo-tube is mounted as close to the cathode ray screen as possible. This photo-tube should have as large a cathode surface as possible since the cathode surface should intercept almost all of the light emitted at the screen. It is also necessary that the fraction of light intercepted be independent of the position of the bright line on the screen. This demands that the cathode area be larger than the screen area used.

The opaque diaphragm shown in Fig. 6 is then interposed between the fluorescent screen and the photo-tube. As a result of the presence of this diaphragm, the amount of light which the photo-tube receives changes as the bright line is swept from left to right. The vertical line AB represents the bright line on the oscillograph screen at one position during its journey across the screen. It should be noted that changes in the length of the line AB are proportional to changes in $1/C$ since amplifiers 1 and 3 are both linear.

The photo-tube and the one-tube amplifier circuit (amplifier 3 of Fig. 2) shown in Fig. 7 have been found to operate satisfactorily. The oscillograph used was a Dumont 208 with a blue short persistance screen. The measured maximum a.c. output voltage furnished by circuit 7 was 0.2 volt which is certainly larger than the 0.08 volt required. The portion of the cathode ray screen used was small, the rectangle being one inch along the horizontal and a quarter inch along the vertical. This variable voltage generator as we used it produced a voltage variation which was the same as the input curve cut in the opaque diaphram to within ±five percent. The accuracy could be increased to two percent if a photo-tube of larger cathode area were used.

In Fig. 6, $1/C$ is kept constant for about half

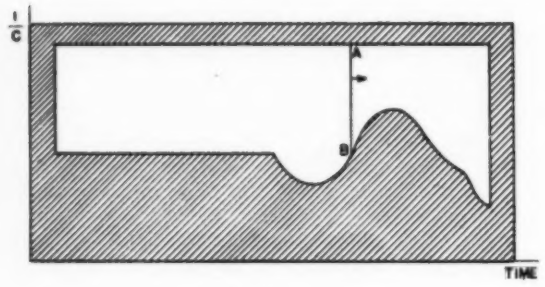


FIG. 6.

the period of the instrument. This time is used to establish the initial conditions. This matter will be dealt with later in the paper.

THE VARIABLE RESISTOR

The variable resistor uses a feedback amplifier as shown in Fig. 2 (amplifier 2). According to H. J. Reich,[1] the equivalent circuit of this feedback amplifier is as given in Fig. 8, where e is the input voltage, ne is the output voltage of amplifier 2, R_2 is the input impedance, and R_3 is the output impedance. The input current i is:

$$i=\frac{e}{R_2}-\frac{(n-1)e}{R_1-(j/\omega C_3)}.$$

If R_2 is very large, then the first term is negligible. If C_3 is very large, the impedance introduced by it at the frequencies used in the resonant circuit is negligible. The effective resistance introduced into the resonant circuit by the feedback amplifier is therefore:

$$\frac{-R_1}{n-1}=\frac{-R_1}{n_0-1}\left\{1-\frac{\Delta}{n_0-1}-\frac{\Delta^2}{(n_0-1)^2}+\cdots\right\}. \quad (7)$$

where $n=n_0+\Delta$ and Δ has been assumed small.

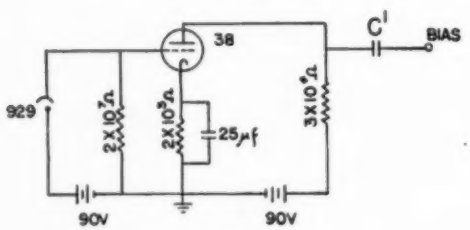


FIG. 7.

[1] H. J. Reich, *The Theory and Applications of Electron Tubes* (McGraw-Hill Book Company, Inc., New York, 1945), 2nd edition, p. 214.

If the additional resistance ρ in the series resonant circuit is adjusted to the value

$$\rho=\frac{R_1}{n_0-1}\left\{1-\frac{\frac{1}{2}\Delta_{\max}}{n_0-1}\right\},$$

then the total resistance R in the series resonant circuit is

$$R=\frac{R_1(\Delta-\frac{1}{2}\Delta_{\max})}{(n_0-1)^2}. \quad (8)$$

The following values for the various parameters were found suitable: $R_2=10^6\Omega$, $C_3=50\ \mu f$, $n_0=1.4$, $\Delta_{\max}=0.016$. When Δ varies from zero to $\Delta_{\max}$, R varies from $-6.25\times10^5\Omega$ to $+6.25\times10^5\Omega$. This range of values makes available suitable time constants. For example, C might discharge through R during charging. R is therefore made large and positive during charging. The smallest discharge time then possible is $R_{\max}C_{\min}=0.312$ second which is long in comparison with the period.

The amplifier 2 used in the variable resistance is shown in Fig. 9. The circuit used to vary the *BIAS* voltage of amplifier 2 is a duplicate of the variable voltage generator previously discussed.

THE CHARGING CIRCUIT

The portion of the circuit labeled *CHARGE* in Fig. 2 is shown in detail in Fig. 10. The very large plate current of the 813 tube is used to charge the condenser C. The square wave generator synchronizes the operations of the various components of the integrator. In this case the square voltage wave introduced at the *BIAS* terminal is amplified and then suffices to turn the plate current on and off in the 813 tube. For large initial values of C (0.085 farad) the average plate current during the quarter cycle used for charging is 0.673 ampere. The voltage across C when charged is 0.066 volt. Battery B of Fig. 2 constitutes the plate supply of the 813 tube. The required supply voltage is 1000 volts.

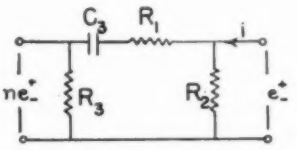


FIG. 8.

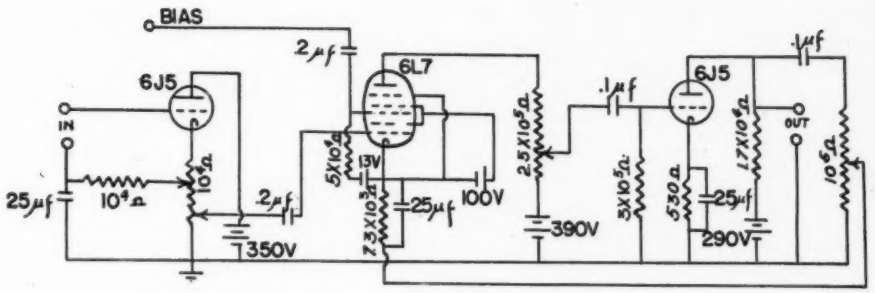


FIG. 9.

THE IMPORTANCE OF AVOIDING FLUCTUATIONS

Fluctuations in the voltages present will produce changes in q. The resulting error introduced into the solution persists throughout the cycle in question. The fact that the curve observed visually is an average over several cycles is of some assistance, but it is also true that the effects produced by several fluctuations which occur during the same cycle can be cumulative. It is important therefore that all power supplies, especially those associated with the first stages of the various amplifiers, be well regulated. In the experiments reported here batteries were used. It is also important to avoid current fluctuations especially in the various vacuum tubes used.

THE ESTABLISHMENT OF THE INITIAL CONDITIONS

During the first quarter of a cycle the condenser C is charged. During the second quarter of a cycle the initial conditions are established. During the last half of the cycle R and C are varied according to the dictates of the differential equation being investigated and the resulting q is traced on the oscilloscope screen. The initial conditions are established as follows: The capacitance of the condenser C is held constant during the first half-cycle at the value desired at the beginning of the second half-cycle (see Fig. 6). At some instant during the second quarter of a cycle, R is changed from infinity to the value desired at the beginning of the second half-cycle. This value of R is maintained until the second half-cycle begins. After R is made finite the circuit obeys the differential equation,

$$L(d^2q/dt^2)+R(dq/dt)+q/C=0, \qquad (9)$$

where LC and R are constant for the remainder of the second quarter of a cycle. Solving Eq. (9), remembering that dq/dt is zero at the time R is reduced to a finite value, one finds that

$$q=\frac{q_0\alpha_2}{\alpha_2-\alpha_1}\left\{e^{\alpha_1 t}-\frac{\alpha_1}{\alpha_2}e^{\alpha_2 t}\right\}, \qquad (10)$$

where t is zero when R is reduced to a finite value and q_0 is the charge on the condenser when t is zero; in addition, α_1 and α_2 are the two values of α given by

$$\alpha=\tfrac{1}{2}\left\{-\frac{R}{L}\pm\left(\frac{R}{L^2}-\frac{4}{LC}\right)^{\frac{1}{2}}\right\}. \qquad (11)$$

The values of q and dq/dt can be made to have the desired values by adjusting q_0 and τ where τ is the time during the second quarter of a cycle when R is finite.

SETTING UP A PROBLEM; THE CHOICE OF THE PROPER TIME CONSTANTS

The way in which the time constants characteristic of the electrical analog of a differential equation are determined can be most easily understood by considering a simple example. Let us therefore discuss how the apparatus would be used to solve the following problem: Given an electron moving in the one-dimensional potential shown in Fig. 11. The Schroedinger wave function ψ for an electron having an energy E of 10 electron volts and having zero slope at x equals zero is to be found. The wave function ψ is an

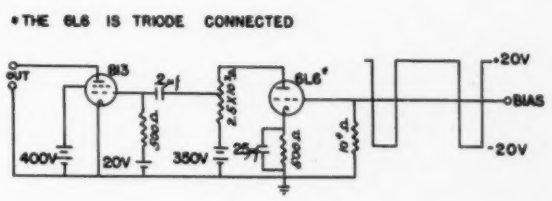


FIG. 10.

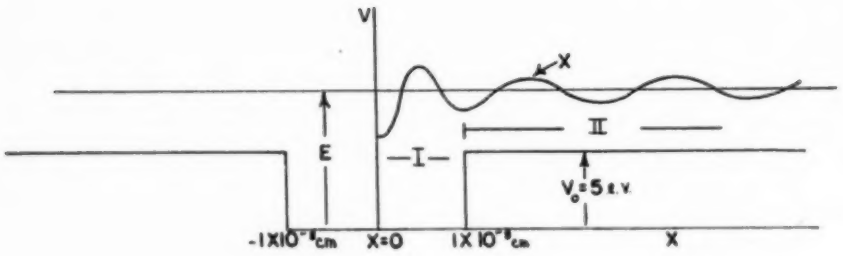


FIG. 11.

even function of x. In region I the wave equation is

$$\frac{d^2\psi}{dx^2}+\frac{8\pi^2m}{h^2}\cdot E\psi=0. \qquad (12)$$

In region II the wave equation is

$$\frac{d^2\psi}{dx^2}+\frac{8\pi^2m}{h^2}\{E-V_0\}\psi=0, \qquad (13)$$

where m is the mass of the electron, h is Planck's constant, and E and V_0 are given in ergs. In Fig. 11, ψ has been drawn to agree with certain qualitative notions concerning the nature of the solution.

The electrical analog of this problem must now be considered. It is clear that only the first quarter of a cycle is needed to establish the initial conditions since it is required that dq/dt be zero at the time t equals zero. The differential equation for the electrical analog is

$$(d^2q/dt^2)+(q/LC)=0, \qquad (14)$$

where $1/LC$ varies with time in the manner shown in Fig. 12.

If the electrical and the wave mechanical solutions are to be comparable, then the "wavelengths" in the two cases must be equal, i.e., there must be as many complete oscillations of q in region I as there are of ψ in region I. This condition will be fulfilled if

$$x_1^2\cdot\left(\frac{d^2\psi/dx^2}{\psi}\right)_I=t_1^2\left(\frac{d^2q/dt^2}{q}\right)_I. \qquad (15)$$

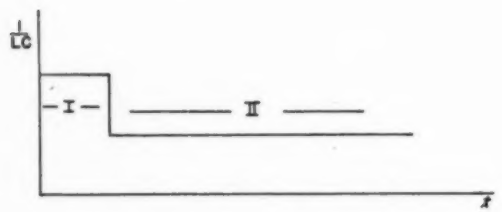


FIG. 12.

Substituting from (12) and (14) this becomes

$$x_1^2(8\pi^2mE/h^2)=t_1^2(1/LC)_1. \qquad (16)$$

If we wish to have region I account for one-fifth of the time during which our solution is being traced and if a third of a period is needed to establish the initial conditions, then $t_1=1/30\times2/3\times1/5\cong0.00444$ second. From (16) $C_1=1.79\times10^{-4}$ farad and if the initial voltage drop across the condenser is 0.066 volt, then q_0, the initial charge on the condenser, is 1.18×10^{-5} coulomb.

Thus, during the first quarter of a cycle the condenser is charged. During the next tenth of a cycle R is kept very large and positive, eliminating any transients which might be produced at the end of the charging period. During the remaining two-thirds of a period R is made zero and the value of $1/LC$ is varied in the manner shown in Fig. 12.

POSSIBLE GENERALIZATIONS

The use of the apparatus in solving linear ordinary differential equations of the second order has been discussed. It is clear that various other problems can be solved using slightly modified apparatus. In the discussion below the type of differential equation being considered is noted, and the necessary modifications in the circuits are pointed out.

1. Differential Equations of Third Order or Higher

Differential equations of third order or higher can be investigated by using two or more series resonant circuits coupled together in various ways. For example, consider the system shown in Fig. 13. The differential equations describing the system are

$$\begin{aligned}&L_1(d^2q_1/dt^2)+R_1(dq_1/dt)+q_1/C_1=0,\\&L_2(d^2q_2/dt^2)+R_2(dq_2/dt)+q_2/C_2=0,\\&L_1(d^2q_1/dt^2)=(1/C_2)q_2.\end{aligned} \qquad (17)$$

Eliminating q_2 from Eqs. (17) the differential equation describing the changes in q_1 with time is

$$L_2\frac{d^2}{dt^2}\left(C_2R_1\cdot\frac{dq_1}{dt}+\frac{C_2}{C_1}q_1\right)+R_2\cdot\frac{d}{dt}\left(C_2R_1\frac{dq_1}{dt}+\frac{C_2}{C_1}q_1\right)+R_1\frac{dq_1}{dt}+\frac{q_1}{C_1}=0. \quad (18)$$

This is a third order differential equation. The three coefficients of importance in such a differential equation can be made to have the desired time dependence by properly varying C_1, R_1, C_2, and R_2 with time. Since there are four quantities which can be varied and only three coefficients, there can be one relation between the coefficients.

It is evident that the coupling can be accomplished in various ways. A fourth-order equation can be obtained by requiring that the voltage across R_1 be equal to the voltage across C_2. For higher order equations three or more coupled circuits are required.

2. Non-Linear and Non-Homogeneous Differential Equations

Non-linear and non-homogeneous differential equations can be investigated using this apparatus. The following example will illustrate the method. Consider the differential equation

$$d^2q/dt^2+q(dq/dt)+q=F(t). \quad (19)$$

The apparatus described will trace a q *versus* t curve which represents a solution of this equation if a voltage source whose voltage varies with time as $F(t)$ is inserted into the series resonant circuit. In addition it would be necessary to hold C constant and vary R in such a way that R increased linearly with the voltage across C.

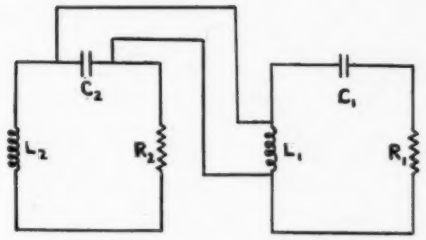


FIG. 13.

3. Boundary Value Problems

One dimensional boundary value problems can be solved by the apparatus. For example, consider the wave mechanical problem previously discussed when E is less than V_0. In this case only those solutions having $\psi=0$ at $X=\infty$ are acceptable, and one can only find solutions satisfying the boundary conditions at $X=0$ and at $X=\infty$ for certain E values. The procedure in the electrical case would be to arrange matters so that at the initial time $dq/dt=0$; then vary C_{I} and C_{II} until the value of q at large values of t goes to zero. It should be noted that a given E value determines both C_{I} and C_{II} so that C_{II} must be changed by an amount consistent with the C_{I} change. Once satisfactory values of C_{I} and C_{II} have been obtained, the energy values E which are allowed can be calculated.

4. Boundary Conditions Not Given for the Initial Time

Suppose we wish a solution of a differential equation which satisfies the boundary conditions $q=a$ at $t=0$ and $q=b$ at $t=T$. The electrical apparatus should be operated in such a way that $q=a$ at $t=0$ and the initial value of dq/dt should be varied until an initial dq/dt is found such that $q=b$ at $t=T$. The curve traced in this case would represent the desired solution.

The Rising Sun Magnetron*

S. MILLMAN** AND A. T. NORDSIECK**
Columbia Radiation Laboratory, Columbia University, New York, New York
(Received July 15, 1947)

A method for separating the π-mode from neighboring resonant modes in a multi-cavity magnetron without the use of "straps" is described. The new resonant circuit design known as the "rising sun" design utilizes two different resonator sizes arranged symmetrically around the cylindrical cathode-anode space, with resonators alternately larger and smaller. Some of the essential features of the mode spectrum and the electromagnetic field in the interaction space in this design are derived from a simple lumped-constant theory. The advantages of the rising sun over the strapped anode block become important in the short wave-length part of the microwave region, because of simpler construction and smaller radiofrequency losses. For very high power magnetrons the advantages of the rising sun structure apply to wave-length as long as 10 cm.

I. INTRODUCTION

THE multi-segment magnetron has in the past few years been developed to considerable perfection for generating very high power electromagnetic oscillations in the decimeter and centimeter wave-length ranges.[1] One of the most important problems in the design of such magnetrons is to insure that the operation is in only one of the many possible resonant frequencies of oscillation of the multi-cavity structure. It has been found that this objective is most easily accomplished if the circuit is designed so that the operating resonant frequency is well separated from all other resonant frequencies of the structure. If a number of identical resonators, say N of them, are formed in a metal block so that they are disposed close together in a ring around a central hole, as for example in Fig. 1a, and if no special measures are taken to couple them together, the N resonant modes of the whole structure tend to be bunched together in frequency near the resonant frequency of an individual resonator. This is a consequence of the small mutual coupling among the individual resonators. If there were no coupling at all, the N resonant modes would all coincide in frequency.

The resonant modes of a multi-segment magnetron anode block are normally characterized either by an integer n, defined as the total electrical phase change around the anode block measured in revolutions; or by the phase difference φ between successive segments. For an anode block with N segments these two quantities are related by the equation

$$\varphi_n = 2\pi n/N. \tag{1}$$

Figure 1a shows the mode spectrum of a symmetrical anode block of 18 segments, accompanied by a sketch of the anode block structure. The spectrum is exhibited as a plot of the free-space wave-length λ, corresponding to the resonant frequency for each value of n. All of the modes except the $n=0$ mode (at infinite wave-length) and the $n=9$ mode are doubly degenerate, so that there are 18 modes in all. For a number of reasons the $n=9$ mode or in general the "π-mode" (mode in which $\varphi=\pi$ or the successive segments are at equal and opposite potentials) is almost invariably chosen for operation. The most important reason is that this mode, which exists only for anode blocks with even N, is not degenerate so that the frequency determines the electromagnetic field pattern uniquely. Now it will be observed that the simple symmetrical structure of Fig. 1a has a particularly poor mode frequency separation at frequencies near the π-mode frequency. The advantages of operation in the π-mode are so considerable that the

* This work is based on work done for the OSRD under Contract OEMsr-485.
** Now at Bell Telephone Laboratories, Inc., 463 West Street, New York, New York.
[1] See numerous abstracts for the Cambridge-American Physical Society meeting of April 25–27, 1946, published in Phys. Rev. 69, 677 (1946). See also a comprehensive article "The magnetron as a generator of centimeter waves" by Fisk, Hagstrum, and Hartman, in the Bell Sys. Tech. J. for April, 1946. We shall refer to this article as reference 1. An abbreviated form of the last articles has been published by Hagstrum, I.R.E. 35, 548 (1947).

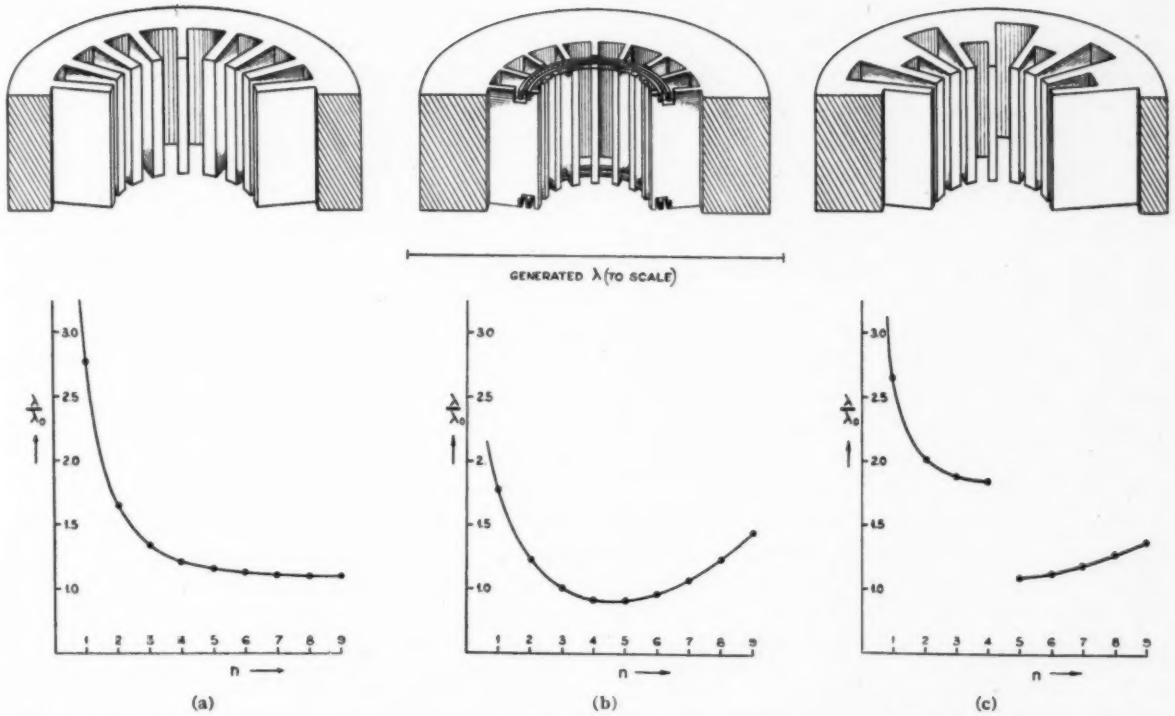


FIG. 1. Three types of microwave magnetron anode blocks: (a) Unstrapped symmetrical anode block, (b) double-ring strapped anode block, (c) rising sun anode block. The relative dimensions are chosen so that each type generates approximately the same free-space wave-length, which is illustrated to scale by the horizontal line segment. Underneath each anode is given a graph of a typical mode frequency spectrum associated with that anode type. The wave-length unit, λ_0, appearing in the ordinates of the graphs is the resonant wave-length of an individual cavity resonator in case (a), of a resonator without strap in (b) and of the smaller resonator in (c). The particular resonator shape illustrated ("vane and sector") is not essential to the discussion in the text, which applies equally to other resonator shapes commonly used.

development of means of separating its resonant wave-length constituted one of the most important problems in making reliable and efficient magnetron oscillators.

Two successful means of separating the π-mode wave-length are shown in Fig. 1b and Fig. 1c.[2] In Fig. 1b is illustrated a recent version of the technique of "strapping," i.e., connecting alternate segments together by means of metallic strips, called straps. This method was invented by British physicists at the University of Birmingham in 1941. All those segments which, in the π-mode, have the same radiofrequency potential, are connected together. If the straps functioned as complete short circuits, we should have only one mode, the desired one, at finite frequency. However, on account of the very high frequencies involved, the straps have an appreciable inductance between connections, so that their effect is as shown in the graph of Fig. 1b.[3] The second means of separating the π-mode frequency, which forms the subject proper of the present article, is shown in Fig. 1c. This means was developed by the authors with the cooperation of many members of the staff of Columbia Radiation Laboratory in 1944.[4] As is apparent from the sketch of the anode structure,

[2] Other means of separating the π-mode besides the two illustrated are also known but these other means have at present a limited field of application. For example, a four- or six-segment symmetrical unstrapped anode block can often be made to function in the π-mode because, although the mode-separating coupling is weak, there are so few modes to separate. This method is thus limited to small numbers of segments. Another example is the so-called interdigital anode block, in which there are no individual resonators proper but a closed-ring transmission line has segments ("digits") attached alternately to opposite conductors of the line. This technique is limited at present to anode blocks in which the ratio of anode circumference to generated wave-length is small. Accordingly, this type of magnetron is suitable for relatively low voltage, low power operation.

[3] A detailed analysis and discussion of strapping is given by L. R. Walker in Chapter 4 of *The Microwave Magnetron* (McGraw-Hill Book Company, Inc., New York, in press).

[4] See National Defense Research Committee, Division 14, Report No. 588, April 1, 1946, Sections 10 and 11. This report is available from the Library of Congress.

there are no straps involved, but the resonators are made alternately larger and smaller than in the symmetrical unstrapped anode for the same frequency.[5] The name "rising sun" has been associated with this structure because of the appearance of the median plane section of such an anode block.[6]

In the following sections we shall describe how the rising sun structure was arrived at (Section II), discuss the properties of this type of structure, and point out the need for a more exact analysis (Section III) as given by the electromagnetic theory of the accompanying paper by Kroll and Lamb,[7] and finally indicate fields of application for which the "rising sun" design is particularly advantageous (Section IV).

One may ask why an alternative to the strapping method was sought. The authors were particularly interested in generating very short wave-lengths, of the order of 1 cm or less, and in this wave-length region the strapping problem takes on watch-making proportions, as the reader may gather from Fig. 1b. We do not mean to imply that strapping is impossible near 1 cm; indeed, very successful 1.25-cm strapped magnetrons were made by Dr. S. Sonkin and one of the authors (A.N.) at Columbia Radiation Laboratory.[8] But the technical difficulties are considerable. Another undesirable feature of the strapped anode block, which also becomes increasingly restrictive in the design of magnetrons at shorter and shorter wave-lengths, is the relatively high radiofrequency copper losses. These losses depend directly on the surface-to-volume ratio of the resonators and are greater for the strapped anode block, as is apparent from a comparison of Figs. 1b and 1c.

II. ATTEMPTS TO IMPROVE NON-π-MODE OPERATION

The initial attack on the problem of producing a reliable unstrapped magnetron for very short wave-lengths took the form of seeking improvements in the behavior of a symmetrical unstrapped magnetron of the type shown in Fig. 1a. Such a magnetron normally operates in an intermediate mode, for which φ is in the neighborhood of 90° rather than 180°. Furthermore, the operation is usually of the "Hartree harmonic" type, as explained in reference 1, pp. 182–192. Thus an 18-segment symmetrical unstrapped magnetron might operate in the $k=-14$ harmonic of the $n=4$ mode, or a 12-segment symmetrical unstrapped magnetron might operate in the $k=-9$ harmonic of the $n=3$ mode. (k and n have the same significance as in reference 1.) One of the factors governing the likelihood of operation in this sort of mode is the fact that the mode separation is appreciable near it so that the electrons have a more determinate radiofrequency field pattern to interact with, whereas at or near the π-mode frequency in such a magnetron, the fields would be an arbitrary mixture of a large number of patterns, resulting in very confused electron orbits and inefficient and unstable operation. The advantage of operating a magnetron in a harmonic of a mode of low n value, provided it could be made reliable and efficient, would be a large size for a given wave-length and voltage.

In spite of considerable efforts, including those about to be described, this non-π-mode type of operation has not been made reliable and efficient.[9] There are probably three main reasons for poor operation in such a mode: (a) The mode is a doublet, so that ideally the frequency does not determine the field pattern uniquely and small asymmetries in the anode block structure due to machining errors or the output coupling cause uncontrolled amounts of doublet splitting. (b) In harmonic operation there is superposed on the harmonic component of the field, which alone is the useful part of the field, a relatively large fundamental component which distorts the electron orbits harmfully. (c) Since the operating value of k is rather large there is a considerable

[5] See also reference 1, pp. 227–232 for a discussion of the rising sun anode structure.

[6] Certain magnetrons developed by the Japanese scientists have a characteristic anode block structure with alternate resonators of different shapes and sizes, very similar to the rising sun structure. The authors have not been able to determine to what extent the principles underlying the alternating unequal cavity design were appreciated by the Japanese workers.

[7] N. M. Kroll and W. F. Lamb, J. App. Phys. 19, 166 (1948).

[8] See Section 20 of report listed in reference 4.

[9] A limited amount of success had been obtained, in the Columbia Radiation Laboratory, with a 14-resonator unstrapped magnetron operating in the mode $k=-11$, $n=3$ when used with a coaxial output, and in the mode $k=-10$, $n=4$ when a wave-guide output was provided. In the latter version a peak power output of over 30 kw was obtained in the 1.25-cm region. See Section 9 of reference 4.

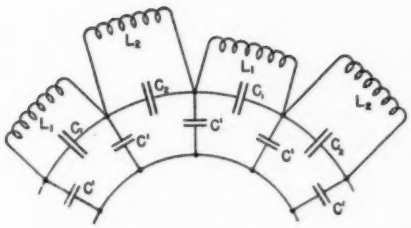


FIG. 2. Lumped-constant circuit used in the theory of the rising sun anode. The ratios $L_1:C_1$ and $L_2:C_2$ are kept equal throughout the discussion. The circuit is supposed to continue around and close on itself and to contain an even number N of resonant circuits. The capacitances C' represent the coupling of the anode segments to one another through the fields in the central hole.

probability that harmonic components of other resonant modes will also satisfy the conditions for oscillation and compete with the intended type of operation.[10]

A way was sought to improve the non-π-mode type of operation by duplicating some of the advantages which were supposed to result from strapping, namely: an increase of mode frequency separation near the desired mode frequency and a distortion of the field patterns of competing modes so as to make them less likely to interact efficiently with electron stream and oscillate in competition with the desired mode.[11] The general scheme for accomplishing these purposes was to introduce cuts or notches in the backs (outer portions) of certain resonators and to determine the resulting mode spectrum and pattern quality by probe measurements. The cuts were merely a convenient way of decreasing the resonant frequency of the chosen resonators by controlled amounts. The placing and arrangement of the cuts was at first dictated by rather vague arguments of symmetry. The cuts were to have the symmetry of the mode to be encouraged and would then, of course, have the wrong symmetry for the adjacent modes and spoil their field patterns.

Several anode block models made along these lines were unsatisfactory because, although the desired mode had a good pattern and the neighboring modes had spoilt patterns, it was invariably found either that the doublet of the desired mode was split or that the improved mode frequency separation came in the wrong part of the spectrum. A simple lumped-constant theory was then worked out to predict the mode spectrum of any ring-shaped structure consisting of two different kinds of resonators following upon one another according to any definite rule. This theory immediately made it clear that the required arrangement was one of large and small resonators alternating in an anode block in which the total number of resonators is twice an odd number. Since the same arrangement is used in the π-mode rising sun anode block to be discussed below, the theory of this arrangement will be presented in detail. (In the π-mode operation the restriction that N be twice an odd number falls away.)

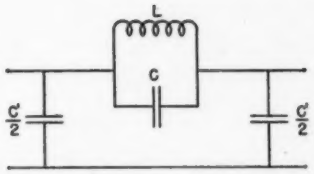


FIG. 3. Elementary section of the circuit of Fig. 2.

The mathematical model chosen was the lumped-constant circuit shown in Fig. 2, the L_1-C_1 circuit representing the smaller (higher frequency) resonator and the L_2-C_2 circuit the larger (lower frequency) resonator. In the special case $L_1=L_2$, $C_1=C_2$ this circuit reproduces the observed mode spectrum of the symmetrical anode quite well and is the simplest circuit which does so. The ratios $L_1:C_1$ and $L_2:C_2$ are kept equal throughout the discussion. This means that the two types of resonators are assumed to have equal characteristic impedances even when their resonant frequencies differ, and corresponds roughly to the situation with the actual cavity resonators. The mutual coupling which gives rise to the mode separation already present in the symmetrical unstrapped anode block is represented by a coupling capacity C' between every "segment" and a common point.

The mode spectrum of the circuit of Fig. 2 can be found by the methods of filter circuit analysis.[12] For the special case $L_1=L_2$, $C_1=C_2=C$,

[10] A detailed discussion of harmonic operation and the competition of harmonic modes is given in an article "Rising sun magnetrons with large numbers of resonators" by Hollenberg, Kroll, and Millman, soon to be published in this journal.

[11] This latter result is accomplished in the strapped anode structure by judiciously introducing gaps in the strap rings ("strap breaks") at points which are current nodes for the π-mode but not for the interfering modes.

[12] See Page and Adams, *Principles of Electricity* (1934), pp. 530–540; or any text on filter circuits.

we have a ring of N sections, each like Fig. 3. This is a so-called pi-network and its phase constant, φ, is given for any wave-length λ by

$$\cos\varphi = 1 + \frac{Z_s}{2Z_p} = 1 + \frac{C'}{2C}\frac{\lambda_0^2}{\lambda_0^2 - \lambda^2}, \tag{2}$$

where Z_s is the impedance of the bar of the pi-section, $2Z_p$ is the impedance of one leg of the pi-section, and λ_0 is the common resonant wave-length of all the cavities. So long as $L_1 = L_2$ and $C_1 = C_2$ each pi-section is terminated by an identical pi-section (i.e., is matched) and therefore the phase shift across each section equals the phase constant. The wave-lengths of the resonant modes of the complete circuit are therefore given by

$$\frac{\lambda}{\lambda_0} = \left(1 + \frac{C'/2C}{1 - \cos\varphi}\right)^{\frac{1}{2}}; \quad \varphi = \frac{2\pi n}{N}. \tag{3}$$

For the general case represented by Fig. 2 with L_1 and L_2 unequal and C_1 and C_2 unequal, we must consider as the fundamental section of our circuit two pi-sections as in Fig. 3, one containing an inductance L_1 and capacitance C_1, the other an inductance L_2 and capacitance C_2, connected to and following upon one another. The whole ring is then made up of $N/2$ such fundamental sections in repetition. For the individual elementary sections we have

$$\cos\varphi_1 = 1 + \frac{C'}{2C_1}\frac{\lambda_1^2}{\lambda_1^2 - \lambda^2}; \qquad \cos\varphi_2 = 1 + \frac{C'}{2C_2}\frac{\lambda_2^2}{\lambda_2^2 - \lambda^2}, \tag{4}$$

where φ_1 and φ_2 are the phase constants of the separate sections containing L_1 and L_2, respectively, and λ_1 and λ_2 are the resonant wave-lengths of the $L_1 - C_1$ and $L_2 - C_2$ circuits, respectively. We need to know the phase difference Φ across the combined fundamental section as a function of the frequency or wave-length. This total phase difference is not the sum of the phase constants φ_1 and φ_2 because neither individual section is operating under matched conditions. Rather Φ must be found by applying Kirchhoff's laws to the fundamental section just as in reference 12. We assume a current and voltage I, V at the beginning of the section and a current and voltage I', V' at the end of the section and express I' and V' in terms of I and V by repeated applications of Kirchhoff's laws. The solution of the equations is then found in the usual way by putting

$$I' = Ie^{-i\Phi}, \quad V' = Ve^{-i\Phi}, \tag{5}$$

and determining the values of Φ and of the ratio

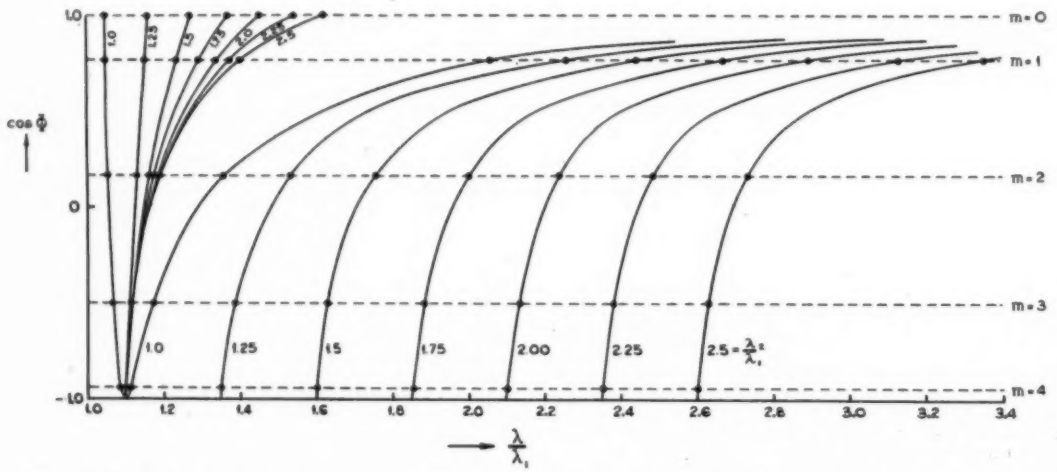


FIG. 4. Graphical derivation of mode spectrum of 18-segment anode with alternating large and small cavities, for several values of the ratio $r = \lambda_2/\lambda_1$ of the resonant wave-lengths of the two cavities. The curves are labelled with the value of r. Each curve except that corresponding to $r = 1$ has two branches. The coupling parameter $C'/2C_1$ is assumed to have the value 0.20. The abscissae of the intersections of the curves with the dotted lines $\cos\Phi = \cos(4\pi m/N)$ give the resonant wave-lengths of the modes in accordance with Eq. (7).

$V:I$ so that the above is a solution. After some calculation we find

$$\cos\Phi = 2\cos\varphi_1 \cos\varphi_2 - 1. \qquad (6)$$

Since φ_1 and φ_2 are given functions of the frequency or wave-length, the above formula gives Φ as a function of the wave-length, and the resonant modes may be found as the roots of the equations:

$$\Phi(\lambda) = \frac{2\pi m}{N/2}; \quad m = 0, 1, 2 \cdots. \qquad (7)$$

Perhaps the best way to exhibit the behavior of the roots is by plotting $\cos\Phi$ against λ/λ_1 for various values of λ_2/λ_1, that is, for various ratios of sizes of the large and small resonators. This is done in Fig. 4 for $C'/2C = 0.20$, a value which is not unusual in practice.

The curves of Fig. 4 have two branches so that there are two resonant modes for each value of m. The assignment of an integer n to each mode is arbitrarily made as follows: For $\lambda_2/\lambda_1 = 1$ (symmetrical case) we have a unique assignment from the theory of the symmetrical anode block (Eq. (1)). As λ_2/λ_1 increases from the value 1, the intersections in Fig. 4 which defined the resonant modes move continuously and keep, by definition, the same value of n. According to this rule $n = m$ on the right-hand branch and $n = N/2 - m$ on the left-hand branch. We must not assume however that for λ_2/λ_1 considerably greater than 1 the electromagnetic field patterns are similar to those for corresponding n values in the symmetrical anode. The question of the field patterns of the modes is touched upon below and is discussed thoroughly by Kroll and Lamb.[7] The nature of the two modes belonging to a given m may be understood qualitatively as follows: Each value of m corresponds to a mode or pattern of the partial resonator system consisting of the $N/2$ small resonators and to a similar pattern on the partial system consisting of the $N/2$ large resonators. These two patterns may be combined in two ways to give a resonant mode of the complete system, just as in the case of a pair of coupled simple circuits. If the two patterns are combined in such a way that adjacent large and small resonators are nearly in phase, we get a mode of the right-hand branch, while a combination with

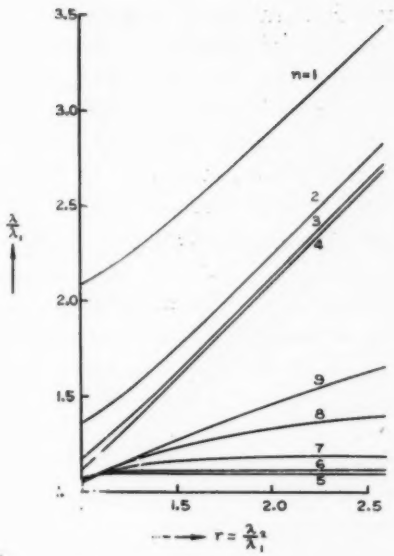


FIG. 5. Resonant mode spectrum of an 18-segment anode with alternating large and small cavities, as a function of the resonator ratio $r = \lambda_2/\lambda_1$. Derived from Fig. 4 by replotting. Note the inversion of order of the short wave-length group of modes at an r value of about 1.1.

adjacent resonators nearly 180° out of phase yields a resonant mode of the left-hand branch. A plot of resonant wave-length *versus* n can be derived from Fig. 4 by reading off the resonant wave-lengths at the appropriate intersections. The graph of Fig. 1c was so constructed for $\lambda_2/\lambda_1 = 1.75$. Another convenient way of presenting information of the type contained in Fig. 4 is to plot the resonant wave-length *versus* the "resonator ratio" λ_2/λ_1 with n as parameter. Figure 5 is derived from Fig. 4 in this way.

The method applied above to the case of large and small resonators alternating can equally well be applied to any regular arrangement of unequal cavities. Indeed the graphical analysis illustrated in Fig. 4 enables us to predict many of the interesting features of the mode spectrum without calculation. Although no arrangement of resonators more complicated than the rising sun arrangement has been found to possess any outstanding advantages, the authors were considerably helped in the earlier stages of this work by the ability to predict the mode distributions of such more complicated arrangements. Suppose that we have a fundamental section consisting of P resonators, all different, and a closed ring of N/P of these fundamental sections. Then we

shall have P branches of the cosΦ *vs.* λ curve, i.e., P groups of modes separated by $P-1$ gaps. Furthermore, if one of the lines cos$\Phi = \cos(2\pi m/(N/P))$ coincides with the line cos$\Phi = -1$, i.e., if N/P is an even integer, the mode or modes $m = N/2P$ will be split doublets. Special cases in which the P resonators of a fundamental section are not all different will show a similar behavior if $C'/2C$ is not too small and the resonators are not too different in size. When the resonators become quite different, the mode spectrum will simplify into a number of groups equal to the number of different sizes of resonator.

The first use made of the above theory was in the attempt to improve non-π-mode operation. The theory made it apparent that we could get the desired increased mode separation about an intermediate mode, without doublet splitting, by choosing alternating large and small resonators and a total number of resonators N equal to twice an odd integer. This choice gives a large separation or gap in the mode spectrum near $n = \frac{1}{2}(N/2-1)$ and does not split the doublet of that mode, or indeed of any mode. As a preliminary to building a magnetron incorporating this idea we constructed a model anode of 14 resonators and followed the mode spectrum and the pattern of the $n = 3$ mode as the large resonators were increased in size step-by-step. The mode spectrum behaved substantially as shown in Fig. 5, except that since there were 14 resonators instead of 18, the gap developed between the 3 and 4 modes instead of between the 4 and 5 modes. At the resonator ratio chosen for operation the nearest neighboring resonant mode to the $n = 3$ mode was the $n = 7$ mode, and it was 25 percent removed in wave-length. The pattern of the $n = 3$ mode was not perfectly regular under these conditions but seemed sufficiently good to justify the expectation of efficient operation in this mode.

Four 1.25-cm magnetrons were built according to this design. All of them operated very poorly, producing at best 2-peak kw output at 2 percent efficiency and changing erratically to oscillation in various modes of the short wave-length group. This behavior was very disappointing and is still not understood completely. The reasons for poor operation in such a mode as listed in the second paragraph of this section apply here in a general way. Reason (c) is even more serious than in a symmetrical anode because the mode patterns of a rising sun anode, with its reduced symmetry, contain extra Hartree harmonic components. Furthermore, although the anode block design was such as to preserve the $n = 3$ doublet unsplit, the output circuits fitted to the four tubes were not well enough designed to eliminate doublet splitting from that source.

III. THE RISING SUN STRUCTURE

Some time after the completion of the experiments described above, attention was focused on the behavior of the mode spectrum of such an alternating large and small cavity structure in the neighborhood of the π-mode, rather than in the region of intermediate n values. It was recognized that when the ratio λ_2/λ_1, denoted by r, is sufficiently large the π-mode region of the spectrum is very similar to the corresponding part of the spectrum of a strapped anode block. The similarity is clear in Fig. 1. The fact that r must be sufficiently large is apparent in Fig. 5, where we see that the short wave-length multiplet has a crossing over or reversal of order at $r \cong 1.1$ and the similarity to the strapped behavior exists only for r well above 1.1. The value of r at the crossing over region depends on the value of C'/C_1 but the general behavior of the mode spectrum is similar for other values of C'/C_1 which occur in practice.

For large enough r one might then expect a magnetron of this type to oscillate in the π-mode. Experience with strapped magnetrons in the frequency range up to 10,000 megacycles/second indicated that the only modes which are dangerous competitors for the π-mode are those of neighboring n, i.e., the $N/2-1$ and $N/2-2$ modes. In regard to these modes the spectrum of the new anode block structure is just like that of the strapped anode block. The first magnetron constructed according to these ideas had 18 resonators, a ratio r of about 1.55 and a separation of 6 percent between the 9 mode and the 8 mode. It oscillated very well and reliably in the 9 mode which was at 24,000 mc/sec. The performance of a typical rising sun magnetron will be described below.

Some additional discussion of the nature of

the rising sun resonant system is perhaps in order here because of the relatively larger r values contemplated and because the emphasis is now on the nature of the π-mode resonance and its field pattern. When $r \geq 2$ and $C'/C_1 \leq .5$ we can get a good qualitative picture of the modes by neglecting C' altogether. The circuit of Fig. 2 with $C'=0$ has three resonant modes which can be found by inspection: There is a resonance at the wave-length where the parallel impedance of the L_1-C_1 sections is infinite, another where the parallel impedance of the L_2-C_2 sections is infinite and a third where the parallel impedances of the two types of sections are opposite and equal. If a small amount of coupling between resonators is introduced the first and second resonances are split up into (in general) $N/2$ closely spaced resonances each, and the third is not split up. These three resonances correspond respectively to the short wave-length group $n=5, 6, 7, 8$, the long wave-length group $n=1, 2, 3, 4$ and the π-mode $n=9$. The π-mode is now regarded as a separate mode not belonging to either group. This classification is consistent with the fact that for large r the π-mode is common to the two sets of resonators. Its resonant wave-length is a weighted average of λ_1 and λ_2 and in this mode the two sets of resonators are approximately equally excited. The short wave-length group of modes is on the other hand almost completely a property of the small resonators, having resonant wave-lengths near λ_1 and excitation almost completely confined to the small resonators; and similarly for the long wave-length group and the large resonators.

These considerations lead one to believe that the electric field pattern of the π-mode will be a good one, i.e., very much like the corresponding pattern of a symmetrical anode block, because the voltages on all the condensers C_1 and C_2 are equal in magnitude with alternating sign in this mode. All the other modes of the rising sun structure, on the contrary, should have poor patterns and a symmetry characterized mainly by the integer m rather than the integer n, because in these modes the structure behaves almost as though it had only $N/2$ resonators. Except for one defect in the π-mode pattern which does not show up in this simple analysis these conclusions are correct. This defect and its consequences are discussed below.

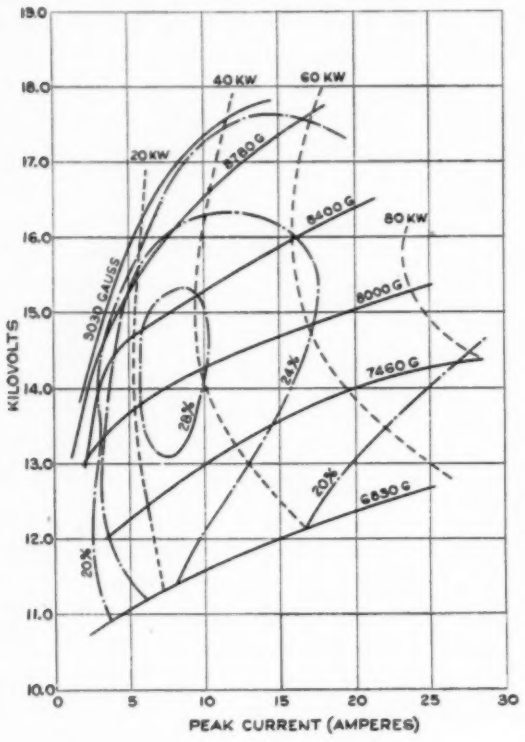
The advantages of this type of anode block structure are considerable and all of them are the more important the shorter the wave-length one wants to generate. It is simpler and has less close spacings and less stringent dimensional tolerances than the strapped anode block for the same wave-length. It can be made in one piece by means of a high precision ground tool by a process called hobbing.[13] The percentage copper losses are smaller than in a strapped anode block for the same wave-length, as was pointed out earlier. Since the percentage copper loss increases as the square root with decreasing wave-length, other factors being comparable, this point is quite important for centimeter and millimeter wave-lengths. In a typical 1.25-cm rising sun magnetron, 35 percent of the radiofrequency power generated is lost in the internal copper, while in a comparable strapped magnetron, 47 percent is lost. A third advantage is the fact that the mode separation and general behavior of the rising sun structure are substantially independent of the height (axial dimension) of the anode block, whereas the strapped anode block runs into mode separation and mode interference difficulties as the height is increased. This means that for the 1-cm range a rising sun magnetron of anode height up to one free space wave-length is practical, while a strapped anode block of height greater than about $\frac{1}{3}$ wave-length is impractical. Finally, the rising sun structure seems to allow larger numbers of resonators. This is because its mode separation is to a considerable degree determined by the ratio r and this parameter can to some extent be adjusted to offset the loss of mode separation resulting from increasing anode diameter.

The performance of a typical 1.25-cm rising sun magnetron is shown in Figs. 6(a) and 6(b).[14] Figure 6(a) shows the lines of constant output power, constant efficiency and constant magnetic field in the anode voltage-anode current plane. Figure 6(b) shows the lines of constant frequency deviation and constant output power in the load

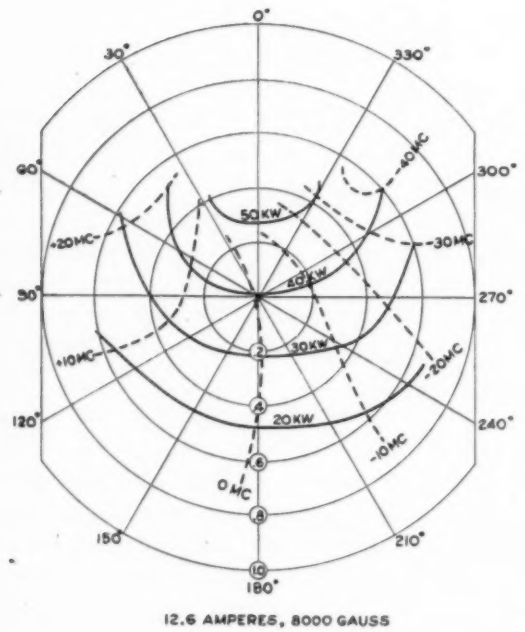
[13] See Section 7 of report listed in reference 4.
[14] These figures correspond to Figs. 13.2 and 13.3 of report listed in reference 4.

impedance plane. This plane is presented in the form of a complex reflection coefficient plane or "Smith chart."[15] This magnetron operates for a time interval of approximately $\frac{1}{2}$ microsecond and repeats about 1000 times per second. The most important dimensions are:

Anode block height	0.190 in.
Anode diameter	0.160 in.
Cathode diameter	0.096 in.
Radial dimension of large resonator	0.115 in.
Radial dimension of small resonator	0.065 in.
Fin thickness	0.017 in.



(a)



(b)

FIG. 6 Performance characteristics of a typical rising sun magnetron for pulsed operation.

Magnetrons of this type have been constructed and operated successfully at other wave-lengths ranging from about 3.4 cm to about 8 mm.[16] Recently similar magnetrons have been constructed and operated at wave-lengths as low as 4 mm.[17]

We now return to the defect in the π-mode pattern of a rising sun anode block. In the simple lumped constant circuit of Fig. 2 the π-mode resonance is characterized by equal and opposite radiofrequency voltages on every neighboring pair of condensers C, and the analogous field pattern of the actual cavity resonator system would be a pure π-mode field. This means that all the fields would vary like $f(N\theta/2)\sin\omega t$ with $f(x+\pi) = -f(x)$. The actual π-mode field pattern is not quite as ideal as this, but corresponds to opposite but not equal voltages on neighboring condensers. The voltage across the gaps of the large resonators is always larger than that across the small resonator gaps in a definite ratio which depends on r and is greater than unity when r is greater than unity. This fact was first brought out by the electromagnetic field treatment of the rising sun resonant structure by Kroll and Lamb.[7] An equivalent way of describing the effect is to say that the field pattern contains some $n=0$ component mixed in with the $n=N/2$ component. The electric field belonging to this additional component consists of circular lines of force in the interaction space. In normal oscillation the electrons interact only with the $n=N/2$ component of the field and the $n=0$ component serves no useful purpose. The $n=0$ component can be a serious source of disturbance

[15] P. H. Smith, "transmission line calculator," Electronics (January 1939).

[16] See Section 14 of report listed in reference 4.

[17] See Columbia Radiation Laboratory Progress Reports for Oct. 1, 1946; Jan. 1, 1947 and March 31, 1947.

for the electron orbits if its amplitude is large and especially if it is in synchronism with the electron motion. This latter condition prevails when the frequency of the cycloidal motion of the electrons is close to the operating frequency. For a plane magnetron synchronism occurs at the cyclotron frequency $\omega = eH/mc$ or at the magnetic field value $H = 10{,}700/\lambda$. In the cylindrical case the synchronous magnetic field is somewhat greater. The practical consequences of the effect just described are as follows: Whereas in a magnetron operating on a pure π-mode field, such as a properly designed strapped magnetron, the efficiency at constant plate current increases monotonically with increasing magnetic field, the rising sun magnetron shows a more or less pronounced dip in efficiency for fields near $12{,}000/\lambda$ gauss. The beginning of this effect is visible in Fig. 6, where the efficiency starts to decrease for fields above 8400 gauss. In other tubes at the same wave-length the observations were carried to higher values of H and the expected increase of efficiency was found. In a 3-cm magnetron where larger values of λH were feasible, the complete course of this effect has been observed, including an efficiency minimum of 28 percent at $\lambda H = 12{,}000$ gauss-cm and a return to normally expected efficiencies of over 50 percent at $\lambda H = 19{,}000$ gauss-cm. The dip in efficiency is deeper and extends over greater ranges of magnetic field the greater the amplitude of the contaminating $n = 0$ component. Hence, a rising sun magnetron should be designed to operate either well above or well below the synchronous magnetic field and with a resonator ratio as small as possible consistent with the requirement of freedom from competing oscillation in the $N/2 - 1$ mode.

In general, the resonator ratio r must be large enough to avoid interfering oscillation in the $N/2 - 1$ mode and small enough to avoid interfering oscillation in a mode of the long wave-length group and disturbance from the zero contamination in the π-mode pattern. These restrictions on r tend to become incompatible for large N. The considerations on the choice of r are discussed in detail in a paper by Hollenberg, Kroll, and Millman.[10]

IV. CONCLUSION

The rising sun magnetron resonant structure is particularly suited for the generation of high power radiation in the short wave-length part of the microwave region. The design permits large cavity resonators, large cathode and anode diameters, and long anodes. These advantages become increasingly important with decreasing wave-length from the points of view of anode block construction, of radiofrequency losses and of thermionic emission and heat dissipation requirements on the cathode. Finally, in the design of very high power (above a megawatt) magnetrons the advantages of the rising sun structure may apply at wave-lengths as long as 10 cm.

ACKNOWLEDGMENTS

This research was carried out under contract OEMsr-485 of OSRD with Columbia University. The authors wish to acknowledge the participation of the Columbia Radiation Laboratory staff in the experiments performed. The reader is referred to reference 4 for a list of the staff, which is too long to reproduce here. The authors are especially grateful for the enthusiastic support of this work by Dr. J. M. B. Kellogg, the Associate Director of the Laboratory.

The Resonant Modes of the Rising Sun and Other Unstrapped Magnetron Anode Blocks*,**

NORMAN M. KROLL AND WILLIS E. LAMB, JR.
Columbia Radiation Laboratory, Department of Physics, Columbia University, New York, New York
(Received July 9, 1947)

In this paper an electromagnetic field theory is developed for unstrapped magnetron anode blocks of rather general symmetry, from which the wave-length spectrum and associated electromagnetic fields can be computed. The method is a generalization of that developed by Clogston (A. M. Clogston, Ph.D. thesis, Massachusetts Institute of Technology, 1941) for the simple case in which all of the side resonators are alike. The characteristics of the rising sun type anode block are discussed in some detail. In addition, the general features of anode blocks having any kind of symmetry are discussed in the limiting case of small anode diameter.

INTRODUCTION

THE magnetron anode block under consideration in this paper consists of a cylindrical structure with a cross section such as appears in Fig. 1. Interest in structures of this type was created by the desire to obtain an unstrapped anode block with a satisfactory distribution of resonant modes. Anode blocks having all resonators identical had proved unsatisfactory, and it was hoped that better results could be obtained using a proper distribution of unequal resonators. The experimental and theoretical work leading to the development of the rising sun structure has been described by Millman and Nordsieck.[1] The calculations described in this paper were begun shortly after the construction of the first rising sun magnetrons and were undertaken in order to obtain detailed information concerning the mode distribution and the corresponding field configurations in the interaction space. In addition, it has been possible to reach some general conclusions concerning the properties of anode blocks having symmetries different from that of the rising sun.

SUMMARY OF THE METHOD

A simplified diagram of the mounting of the anode block appears in Fig. 2. The interior of the magnetron (anode block, end spaces, etc.) can be thought of as constituting a cavity resonator of rather complicated geometry. It is a well-known fact that such a resonator has an infinite number of resonant modes, each characterized by a particular electromagnetic field distribution and frequency of oscillation. The rigorous method of solving this problem is to find continuous solutions of Maxwell's equations which behave properly at the boundaries. However, for a cavity having the complicated geometry shown it is, in general, possible to find only approximate solutions. For the modes of interest in magnetron operation the fields in the anode block are substantially free of axial variation except near the top and bottom of the block. Furthermore, the axial component of the electric field is very small. It is thus reasonable to assume as an approximation that the electric field is transverse and has no axial variation. Under these conditions all field components can be expressed in terms of the axial component of magnetic field, which must satisfy the equation $\nabla^2 H_z + k^2 H_z = 0$ and appropriate boundary conditions. This problem has solutions for certain discrete values of k, and since $k = 2\pi\nu/c$, these determine the resonant frequencies. Thus the three-dimensional vector problem normally involved in cavity resonators is reduced to a two-dimensional scalar one.[2]

* Publication assisted by the Ernest Kempton Adams Fund for Physical Research of Columbia University, New York.
** Initial work supported by OSRD Contract OEMsr-485 and continued with the support of the Signal Corps under Contract No. W36-039 Sc-32003. Some of the results in this paper were reported by the authors in "The resonant modes of the rising sun (A tube) anode," NDRC Report, No. 323, October 25, 1944.
[1] S. Millman and A. Nordsieck, J. App. Phys. **19**, 156 (1948).

[2] J. A. Stratton, *Electromagnetic Theory* (McGraw-Hill Book Company, Inc., New York, 1941), Chapter VI, pp. 349–355. As a matter of fact it is merely necessary to ignore the end spaces to reduce the problem to a two-dimensional scalar one, inasmuch as the anode block itself is cylindrical.

One has now to solve the problem $\nabla^2 H_z + k^2 H_z = 0$ with the boundary condition that the normal derivative of H_z be zero at the metal surfaces.[3] The problem is still too complicated to solve exactly. The approximation used in this paper consists of assuming that the amplitude of the transverse electric field (E_ϕ) is constant across each junction of a side resonator with the interaction space. The time dependence of all fields is of course taken to be harmonic. One can then compute, analytically in most cases, the magnetic field throughout the anode block. It is then possible to find frequencies such that the average values of the magnetic field match at each junction, and these are the resonant frequencies. It is not possible to match the magnetic field exactly at the junctions since the assumed electric field configuration is not correct.[4] The assumption concerning the distribution of electric field further restricts the solution to those modes for which the field configuration is approximated by that assumed. For example, modes for which E_ϕ alternates in sign several times across a resonator opening will not appear in the solution.

THE SYMMETRIC ANODE BLOCK

An anode block in which all of the side resonators are alike is generally referred to as symmetric. The general theory of the resonances of such an anode block has been given by Clogston in his Ph.D. thesis, and by others. Since these discussions have not been published in generally available form, and further since an understanding of the resonances of the symmetric anode block is essential in understanding those of the rising sun and other types of anode blocks, it appears that repetition of the discussion here is justified.

A cross section of the anode to be discussed appears in Fig. 3. In the initial discussion rectangular side resonators will be used. The discussion can readily be generalized to include other shapes of side resonators. The problem of the side resonators and the cathode-anode space will be treated separately. In Fig. 4 there appears a diagram of the q'th side resonator together with a Cartesian coordinate system. A solution of Maxwell's equations satisfying the following conditions is required.

1. The time dependence of all fields must be described by a common factor $e^{j\omega t}$.
2. $E_z = 0$ (i.e., the electric field is transverse).
3. All fields are independent of z.
4. $E_x = 0$ at $y = \pm d/2$, $E_y = 0$ at $x = L$ (i.e., the tangential electric field vanishes at the metal surfaces).
5. $E_y = E_q e^{j\omega t}$ at $x = 0$ (henceforth the factor $e^{j\omega t}$ will be omitted) where E_q is a constant for each resonator.

From conditions 1, 2, and 3 it follows directly from Maxwell's equations that

$$\frac{\partial^2 H_z}{\partial x^2} + \frac{\partial^2 H_z}{\partial y^2} + k^2 H_z = 0, \quad k = \frac{2\pi}{\lambda} = \omega(\mu_0 \epsilon_0)^{\frac{1}{2}}, \tag{1}$$

$$E_x = -(j/k)(\mu_0/\epsilon_0)^{\frac{1}{2}} \frac{\partial H_z}{\partial y}; \qquad E_y = (j/k)(\mu_0/\epsilon_0)^{\frac{1}{2}} \frac{\partial H_z}{\partial x}, \tag{2}$$

$$H_x = 0; \quad H_y = 0.$$

Furthermore, any solution satisfying the above conditions is unique. One can readily verify that the field distribution given by

$$E_y = \frac{E_q \sin k(L-x)}{\sin kL}, \tag{3}$$

$$H_z = \frac{-jE_q \cos k(L-x)}{(\mu_0/\epsilon_0)^{\frac{1}{2}} \cdot \sin kL}, \tag{4}$$

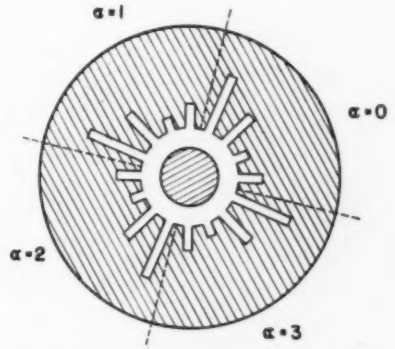


FIG. 1. Cross section of magnetron anode block with unequal side resonators.

[3] Perfectly conducting walls have been tacitly assumed here. For the computation of frequencies and field configuration this approximation is excellent.

[4] Some justification for this procedure appears in the Appendix.

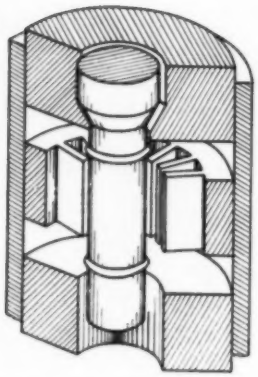


FIG. 2. Cutaway drawing showing the interior of a magnetron.

satisfies all of the above conditions and is therefore the required solution. Thus the magnetic field at $x=0$ is independent of y and is given by

$$H_z=\frac{-jE_q}{(\mu_0/\epsilon_0)^{\frac{1}{2}}}\cot kL. \tag{5}$$

It can be assumed without any loss of generality that fields from slot to slot differ only by a constant phase factor (see Appendix II). Thus E_q can be written as $Ee^{j\Gamma q}$. Furthermore, since $e^{j\Gamma q}$ must equal $e^{j\Gamma(q+N)}$ in order to have each slot in phase with itself, $\Gamma=2\pi n/N$, where n is any integer, and $E_q=e^{(2\pi jn/N)q}$.

A diagram of the interaction space appears in Fig. 5. For this region one requires a solution of Maxwell's equations which satisfies in addition to conditions 1, 2, 3, the boundary conditions (using cylindrical coordinates ρ, ϕ, z)

$$E_\phi(c,\phi)=0$$

and

$$E_\phi(a,\phi)=Ee^{(2\pi jn/N)q} \text{ for } \frac{2\pi q}{N}+\theta\geq\phi\geq\frac{2\pi q}{N}-\theta$$
$$=0 \text{ otherwise.}$$

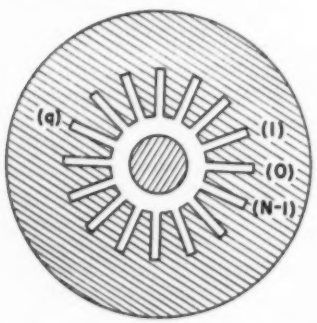


FIG. 3. Symmetric anode block with rectangular side resonators.

The value chosen for $E_\phi(a,\phi)$ insures that the electric field component tangent to the boundary between the side resonators and the interaction space is practically continuous. It would be exactly continuous if the boundary used in discussing the side resonators (i.e., the straight line $x=0$) coincided exactly with the boundary used in discussing the interaction space (i.e., the arc of the circle $\rho=a$ which subtends the resonator opening). As long as the angle subtended by the resonator opening is small, the error incurred is negligible.[5]

From Maxwell's equations and conditions 1, 2, and 3 it follows directly that

$$\frac{1}{\rho}\frac{\partial}{\partial\rho}\left(\rho\frac{\partial H_z}{\partial\rho}\right)+\frac{1}{\rho^2}\frac{\partial^2 H_z}{\partial\phi^2}+k^2H_z=0;$$
$$H_\rho=H_z=0, \tag{6}$$

$$E_\rho=-(j/k\rho)(\mu_0/\epsilon_0)^{\frac{1}{2}}\frac{\partial H_z}{\partial\phi};$$
$$E_\phi=(j/k)(\mu_0/\epsilon_0)^{\frac{1}{2}}\frac{\partial H_z}{\partial\rho}. \tag{7}$$

The fundamental set of solutions for Eq. (6) are the functions $J_\gamma(k\rho)e^{j\gamma\phi}$ and $N_\gamma(k\rho)e^{j\gamma\phi}$, where J_γ and N_γ consist of the well-known Bessel and Neumann functions of order γ.[6] In order that the fields be single valued in ϕ it is necessary that γ be a positive or negative integer or zero. Consider now the set of fields

$$H_z^{(\gamma)}=\left(J_\gamma(k\rho)-\frac{J_\gamma'(kc)}{N_\gamma'(kc)}N_\gamma(k\rho)\right)e^{j\gamma\phi}$$
$$=Z_\gamma(k\rho)e^{j\gamma\phi},$$
$$E_\rho^{(\gamma)}=(\gamma/k\rho)(\mu_0/\epsilon_0)^{\frac{1}{2}}Z_\gamma(k\rho)e^{j\gamma\phi}, \tag{8}$$
$$E_\phi^{(\gamma)}=j(\mu_0/\epsilon_0)^{\frac{1}{2}}\left(J_\gamma'(k\rho)-\frac{J_\gamma'(kc)}{N_\gamma'(kc)}N_\gamma'(k\rho)\right)e^{j\gamma\phi}$$
$$=j(\mu_0/\epsilon_0)^{\frac{1}{2}}Z_\gamma'(k\rho)e^{j\gamma\phi}.$$

[5] In occasional cases for which the angle subtended is large, it is possible to avoid this approximation as follows. The electric field throughout the side resonators has been computed. Inasmuch as the arc of the circle $\rho=a$, which subtends the resonator opening, is contained in the resonator, one can choose the distribution of electric field tangent to this arc to equal that computed for the side resonator.

[6] See, for example, J. A. Stratton *Electromagnetic Theory* (McGraw-Hill Book Company, Inc., New York, 1941), pp. 351–360.

Any linear combination of the fields of the above set will be a solution of Maxwell's equations and will also satisfy the boundary condition at the cathode that $E_\phi=0$ at $\rho=c$, since $Z_\gamma'(kc)=0$. One can now find the linear combination of the functions $E_\phi^{(\gamma)}(\rho, \phi)$ which satisfies the boundary condition at $\rho=a$ by making a Fourier expansion thus,

$$E_\phi(\rho, \phi)=j(\mu_0/\epsilon_0)^{\frac{1}{2}} \sum_{\gamma=-\infty}^{\infty} C_\gamma Z_\gamma'(k\rho)e^{j\gamma\phi}.$$

The constants C_γ are determined by

$$E_\phi(a, \phi)=j(\mu_0/\epsilon_0)^{\frac{1}{2}} \sum_{\gamma=-\infty}^{\infty} C_\gamma Z_\gamma'(ka)e^{j\gamma\phi}=f(\phi),$$

where

$$f(\phi)=Ee^{j2\pi nq/N}$$

for

$$\frac{2\pi q}{N}-\theta<\phi<\frac{2\pi q}{N}+\theta$$

and $f(\phi)=0$ for all other values of ϕ. Then

$$j(\mu_0/\epsilon_0)^{\frac{1}{2}}C_\gamma Z_\gamma'(ka)$$

$$=\frac{1}{2\pi}\int_0^{2\pi} e^{-j\phi\gamma}f(\phi)d\phi$$

$$=(E/2\pi)\sum_{q=0}^{N-1} e^{j2\pi nq/N}\int_{(2\pi q/N)-\theta}^{(2\pi q/N)+\theta} e^{-j\gamma\phi}d\phi$$

$$=E\frac{N\theta}{\pi}\frac{\sin\gamma\theta}{\gamma\theta} \text{ for } \gamma=n+mN, \text{ where } m \text{ is any integer}$$

$$=0 \text{ otherwise.}$$

Thus the fields are given by

$$\left.\begin{aligned} E_\phi(\rho, \phi)&=E\frac{N\theta}{\pi}\sum_{m=-\infty}^{\infty}\left(\frac{\sin\gamma\theta}{\gamma\theta}\right)\frac{Z_\gamma'(k\rho)}{Z_\gamma'(ka)}e^{j\gamma\phi},\\ E_\rho(\rho, \phi)&=-jE\frac{N\theta}{\pi k\rho}\sum_{m=-\infty}^{\infty}\gamma\left(\frac{\sin\gamma\theta}{\gamma\theta}\right)\frac{Z_\gamma(k\rho)}{Z_\gamma'(ka)}e^{j\gamma\phi},\\ H_z(\rho, \phi)&=-j(\epsilon_0/\mu_0)^{\frac{1}{2}}E\frac{N\theta}{\pi}\\ &\quad\times\sum_{m=-\infty}^{\infty}\left(\frac{\sin\gamma\theta}{\gamma\theta}\right)\frac{Z_\gamma(k\rho)}{Z_\gamma'(ka)}e^{j\gamma\phi}\end{aligned}\right\}\quad \gamma=n+mN. \quad (9)$$

The average value of H_z at the q'th resonator opening is given

$$\bar{H}_z=\frac{1}{2\theta}\int_{(2\pi q/N)-\theta}^{(2\pi q/N)+\theta} H_z(a, \phi)d\phi$$

$$=-j(\epsilon_0/\mu_0)^{\frac{1}{2}}E\frac{N\theta}{\pi}$$

$$\times\left(\sum_{m=-\infty}^{\infty}\left(\frac{\sin\gamma\theta}{\gamma\theta}\right)^2\frac{Z_\gamma(ka)}{Z_\gamma'(ka)}\right)e^{2\pi jnq/N}. \quad (10)$$

Resonance is then determined by the requirement that the average value of the magnetic field at each resonator opening as determined by the fields computed for the side resonator match that as determined by the fields computed for the interaction space; thus

$$-j(\epsilon_0/\mu_0)^{\frac{1}{2}}E\frac{N\theta}{\pi}\left(\sum_{m=-\infty}^{\infty}\left(\frac{\sin\gamma\theta}{\gamma\theta}\right)^2\frac{Z_\gamma(ka)}{Z_\gamma'(ka)}\right)e^{2\pi jnq/N}$$

$$=-j(\epsilon_0/\mu_0)^{\frac{1}{2}}Ee^{2\pi jnq/N}\cot kL,$$

or

$$\frac{N\theta}{\pi}\sum_{m=-\infty}^{\infty}\left(\frac{\sin\gamma\theta}{\gamma\theta}\right)^2\frac{Z_\gamma(ka)}{Z_\gamma'(ka)}=\cot kL. \quad (11)$$

It is evident from the fact that terms involving q drop out that satisfaction of the condition at

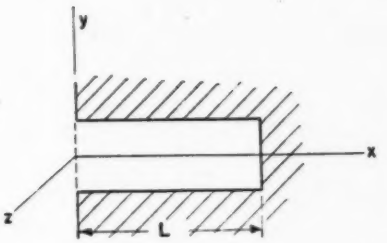


FIG. 4. Rectangular side resonator.

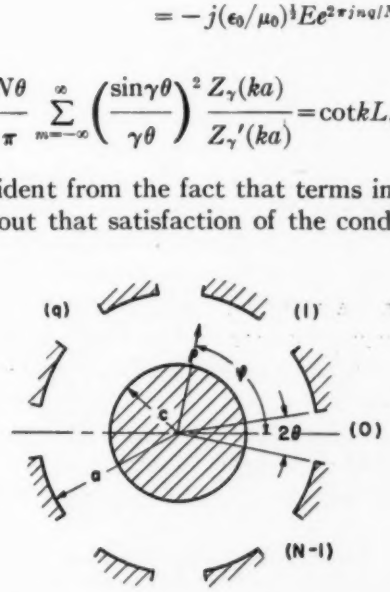


FIG. 5. The interaction space.

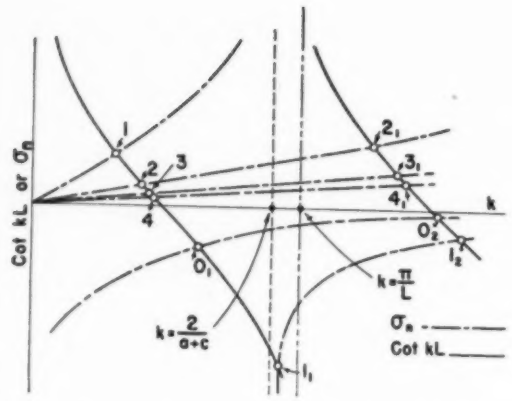


FIG. 6. Graphical representation the equation $\cot kl=\sigma_n$, where

$$\sigma_n=\frac{N\theta}{\pi}\sum_{m=-\infty}^{\infty}\left(\frac{\sin\gamma\theta}{\gamma\theta}\right)^2\frac{Z_\gamma(ka)}{Z_\gamma'(ka)}$$

with $\gamma=n+mN$.

one resonator opening insures that it be satisfied at all of the others. This is, of course, a consequence of the assumption regarding the variation of amplitude from resonator to resonator. Because the electric field tangential to the boundary between side resonator and interaction space has been chosen incorrectly, the magnetic field cannot be made continuous at the junction. In addition, the component of electric field normal to the boundary is also discontinuous (i.e., $E\rho\neq 0$ while $E_z=0$).

Equation (11) is most readily represented graphically as shown in Fig. 6. In addition the general behavior of the functions appearing on the right side of Eq. (11) and denoted $\sigma(n)$ is shown. Inasmuch as the $\sigma(n)$ have the property that $\sigma(n)=\sigma(n+mN)$ and $\sigma(n)=\sigma(-n)$, only the values 0, 1, $\cdots$, $N/2$ or $(N-1)/2$ need be considered for n.[7] There are, however, for each value of n an infinite number of solutions each corresponding to a mode of oscillation. The first roots are the most important, and the associated mode is usually designated by the number n. The modes associated with the higher roots will be designated by n_p (for, say, the $p+1$ root of the equation n). A typical spectrum is shown in Fig. 7 and a table of frequencies of the first group of resonances for a typical case in Table I. All modes are doublets, corresponding to positive and negative values of n, with the exception of those for which $n=0$ or $N/2$. The most significant feature of the spectrum is the small separation (Table I) between the modes of large n. This separation can be increased by increasing the anode diameter and by modifying the shapes of the resonators. Practically speaking, however, it is difficult to obtain a sufficiently well separated π mode ($n=N/2$) for $N>8$.

The distribution of electric and magnetic field in the interaction space is contained in Eq. (9). It consists of a set of harmonic components whose angular variation is given by $e^{j\gamma\phi}$ where γ can take on the values $n+mN$. These are frequently referred to as the "Hartree harmonics." In general, the amplitude of these harmonics falls off as one proceeds from anode to cathode. The rate of decrease increases rapidly with γ, so that near the cathode the harmonic of lowest γ is generally predominant.

In the preceding discussion the side resonators have been rectangular slots. Actually, the rectangular slot is rarely used in magnetrons, the annular sector and the hole and slot being much more common. The previous theory can, however, be readily modified to include any of these cases. The side resonator affects the resonant frequency and interaction field only through the ratio of the magnetic field averaged across the resonator opening to the electric field across the resonator opening. In the case of the rectangular slot this ratio is given by $-j(\epsilon_0/\mu_0)^{\frac{1}{2}}\cot kL$. For any other shape resonator one can define an equivalent length by the relationship

$$\frac{\bar{H}}{E}=-j(\epsilon_0/\mu_0)^{\frac{1}{2}}\cot kL(eq), \qquad (12)$$

where $L(eq)$ is the length rectangular slot resonator which gives the same value of H/E as the resonator in question. $L(eq)$ is, in general, a function of frequency and for many important cases can be found analytically. Thus in the preceding discussion and in the discussion that follows the case of non-rectangular side resonators can be treated simply by replacing the slot lengths with equivalent slot lengths.[8]

[7] This is just what one would expect since replacing n by $n+mN$ has no physical significance, while replacing n by $-n$ merely reverses the direction of the rotating waves, which, of course, should not affect the frequency.

[8] One can also regard the ratio $\bar{H}/E$ as an admittance (perhaps with a geometrical factor in addition) and compute admittances for the various resonator shapes. Ad-

GENERAL THEORY OF UNSTRAPPED ANODE BLOCKS

1. The Resonant Frequencies

A cross section of the anode block to be discussed appears in Fig. 1. There are a total of N resonators (numbered $q=0, 1, \cdots, N-1$) with evenly spaced openings into the interaction space There are N/p identical sections (numbered $\alpha=0, 1, \cdots, N/p-1$) each containing p resonators (numbered $\beta=0, 1, \cdots, p-1$), while the geometries of the various resonators within a section may differ. For the anode block shown in Fig. 1; $N=16$, $p=4$, and $N/p=4$.

The ϕ component of the electric field is assumed to be zero along the metal surfaces of the cathode and anode and constant across each resonator opening at the anode surface. That is,

$$E_\phi(c, \phi)=0,$$

$$E_\phi(a, \phi)=E_q, \text{ for } \left(\frac{2\pi q}{N}-\theta_q\right)<\phi<\left(\frac{2\pi q}{N}+\theta_q\right)$$

$$=0, \text{ otherwise.}$$

It can be assumed without any loss of generality that fields from section to section differ only by a constant phase factor[9] and thus that $E_{q+\alpha p}=E_q e^{j\Gamma\alpha}$. To insure single valuedness one must have $(N/p)\Gamma=2\pi n$ where n is a positive or negative integer or zero, yielding $\Gamma=2\pi np/N$. One can thus conveniently write E_q as

$$(A_\beta e^{(2\pi jn/N)\beta})e^{2\pi jnp\alpha/N}$$

or more simply

$$E_q=A_\beta e^{2\pi jnq/N}, \tag{13}$$

since $q=\beta+\alpha p$. In the case of the side resonators the treatment is identical to that of the preceding section. If these are rectangular as shown in Fig. 1, then the electric field within the resonator is in the y direction (Fig. 4) and given by

$$E_y=E_q\frac{\sin k(L_q-x)}{\sin kL_q},$$

mittance formulae for various shapes appear in volume 6 of the Radiation Laboratory Technical Series, "Microwave magnetrons."

[9] Appendix II.

TABLE I. The first group of resonance for a typical symmetric anode block with $N=18$; $c/a=.6$; $2a/\lambda_0=.325$; $\theta=.068$ radians.

n	λ_n/λ_0	n	λ_n/λ_0
1	1.840	6	1.016
2	1.226	7	1.007
3	1.104	8	1.003
4	1.055	9	1.000
5	1.030		

and the magnetic field, which is axial, is given by

$$H_z=-jE_q(\epsilon_0/\mu_0)^{\frac{1}{2}}\frac{\cos k(L_q-x)}{\sin kL_q},$$

which gives for the magnetic field at the resonator opening

$$H_z=-jE_q(\epsilon_0/\mu_0)^{\frac{1}{2}}\cot kL_q.$$

Then for the case in which the resonators are not rectangular one can represent the magnetic field averaged across the resonator opening by (12)

$$\bar{H}_z=-jE_q(\epsilon_0/\mu_0)^{\frac{1}{2}}\cot kL_q(eq),$$

where it is to be understood that the equivalent length, $L(eq)$, is frequency dependent. In order to compute the magnetic field on the interaction space side of the junctions one computes (with the aid of a Fourier expansion) the electric and magnetic fields throughout the interaction space. Following the procedure of the preceding section one takes

$$E_\phi=j(\mu_0/\epsilon_0)^{\frac{1}{2}}\sum_{\gamma=-\infty}^{\infty}C_\gamma Z_\gamma'(k\rho)e^{j\gamma\phi},$$

$$E_\rho=\frac{(\mu_0/\epsilon_0)^{\frac{1}{2}}}{k\rho}\sum_{\gamma=-\infty}^{\infty}\gamma C_\gamma Z_\gamma(k\rho)e^{j\gamma\phi},$$

$$H_z=\sum_{\gamma=-\infty}^{\infty}C_\gamma Z_\gamma(k\rho)e^{j\gamma\phi}.$$

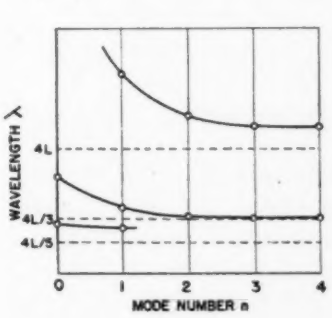


FIG. 7. The spectrum of a typical symmetric anode block with $N=8$.

The constant C_γ are determined by

$$E_\phi(a,\phi)=j(\mu_0/\epsilon_0)^{\frac{1}{2}}\sum_{\gamma=-\infty}^{\infty}C_\gamma Z_\gamma'(ka)e^{j\gamma\phi}=f(\phi),$$

where

$$f(\phi)=A_\beta e^{2\pi jnq/N}\text{ for }\left(\frac{2\pi q}{N}-\theta_q\right)<\phi<\left(\frac{2\pi q}{N}+\theta_q\right)$$

$$=0,\text{ otherwise;}\qquad \beta=q-\alpha p$$

thus

$$j(\mu_0/\epsilon_0)^{\frac{1}{2}}C_\gamma Z_\gamma'(ka)$$

$$=\frac{1}{2\pi}\int_0^{2\pi}e^{-j\gamma\phi}f(\phi)d\phi$$

$$=\frac{1}{2\pi}\sum_{q=0}^{N-1}\int_{(2\pi q/N)-\theta_q}^{(2\pi q/N)+\theta_q}A_\beta e^{2\pi jnq/N}e^{-j\gamma\phi}d\phi$$

$$=\frac{1}{\pi}\sum_{q=0}^{N-1}A_\beta\theta_\beta\frac{\sin\gamma\theta_\beta}{\gamma\theta_\beta}e^{2\pi jnq/N}e^{-2\pi j\gamma q/N}$$

$$=\frac{1}{\pi}\sum_{\beta=0}^{p-1}A_\beta\theta_\beta\frac{\sin\gamma\theta_\beta}{\gamma\theta_\beta}\sum_{\alpha=0}^{(N/p)-1}e^{-2\pi j(\gamma-n)(\beta+\alpha p)/N}.$$

Now

$$\sum_{\alpha=0}^{(N/p)-1}e^{-2\pi j(\gamma-n)(\beta+\alpha p)/N}=\frac{N}{p}e^{-2\pi jm\beta/p},$$

$$\text{for }\gamma-n=mN/p$$

$$=0,\text{ otherwise}$$

where m is a positive or negative integer or zero. Thus

$$C_\gamma=-j(\epsilon_0/\mu_0)^{\frac{1}{2}}\frac{N}{p\pi}\frac{1}{Z_\gamma'(ka)}\sum_{\beta=0}^{p-1}\theta_\beta A_\beta\frac{\sin\gamma\theta_\beta}{\gamma\theta_\beta}e^{-2\pi jm\beta/p}$$

$$\text{for }\gamma=n+mN/p,\quad(14)$$

$C_\gamma=0$, for $\gamma\neq n+mN/p$.

The coefficients C_γ, and consequently the fields in the interaction space, are still not completely determined since the field amplitudes A_β are unknown. These, as well as the resonant frequency, will be determined by the matching conditions at the resonator openings. The average value of the magnetic field at the q'th resonator opening is given by

$$\bar{H}_z=\frac{1}{2\theta_q}\sum_{\gamma=-\infty}^{\infty}\int_{(2\pi q/N)-\theta_q}^{(2\pi q/N)+\theta_q}C_\gamma Z_\gamma(ka)e^{j\gamma\phi}d\phi$$

$$=\sum_{\gamma=-\infty}^{\infty}e^{2\pi j\gamma q/N}\frac{\sin\gamma\theta_q}{\gamma\theta_q}C_\gamma Z_\gamma(ka)$$

which, on inserting the expressions for C_γ, yields

$$\bar{H}_z=-j(\epsilon_0/\mu_0)^{\frac{1}{2}}e^{2\pi jnq/N}\sum_{\beta=0}^{p-1}A_\beta\sigma_{\beta q}(n),\quad(15)$$

where the functions $\sigma_{\beta q}(n)$ are defined by

$$\sigma_{\beta q}(n)=\frac{N\theta_\beta}{p\pi}\sum_{m=-\infty}^{\infty}\left(\frac{\sin\gamma\theta_\beta}{\gamma\theta_\beta}\right)\left(\frac{\sin\gamma\theta_q}{\gamma\theta_q}\right)\quad(16)$$

$$\times e^{-2\pi jm(\beta-q)/p}\frac{Z_\gamma(ka)}{Z_\gamma'(ka)},$$

with $\gamma=n+mN/p$. Resonance is determined as in the symmetric case by requiring that the average values of the magnetic field, as computed on both sides of the resonator openings, match at each opening. Thus

$$-j(\epsilon_0/\mu_0)^{\frac{1}{2}}e^{2\pi jnq/N}\sum_{\beta=0}^{p-1}A_\beta\sigma_{\beta q}(n)$$

$$=-j(\epsilon_0/\mu_0)^{\frac{1}{2}}E_q\cot kL_q,\quad(17)$$

with q set equal to $\beta'+\alpha p$, Eq. (17) reduces to

$$\sum_{\beta=0}^{p-1}A_\beta\sigma_{\beta\beta'}(n)=A_{\beta'}\cot kL_{\beta'}\quad(18)$$

since

$$E_q=A_{\beta'}e^{2\pi jnq/N}\quad\text{and}\quad\sigma_{\beta q}(n)=\sigma_{\beta(q+\alpha p)}(n).$$

β' is used rather than β to distinguish this quantity from the summation index. It is thus apparent that it is sufficient to apply the matching condition at the resonator openings of a single section; matching at the other resonator openings then follows automatically.

Equation (18) represents a system of p linear homogeneous equations in p unknowns (the A_β). Non-trivial solutions for such systems exist only when the determinant of the coefficients is zero.

That is

$$\begin{vmatrix} \sigma_{00}(n)-\cot kL_0 & \sigma_{10} & \cdots & \sigma_{p-1,0}(n) \\ \sigma_{01}(n) & \sigma_{11}(n)-\cot kL_1 & \cdots & \sigma_{p-1,1}(n) \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \sigma_{0,p-1}(n) & \sigma_{1,p-1}(n) & \cdots & \sigma_{p-1,p-1}(n)-\cot kL_{p-1} \end{vmatrix} = 0. \tag{19}$$

Since the coefficients are all functions of frequency, Eq. (19) determines the resonant frequencies.

Certain properties of the determinant in Eq. (19) are of interest. These can be derived from the following easily verified properties of the $\sigma_{\beta\beta'}(n)$.

$$\theta_{\beta'}\sigma_{\beta\beta'}(n)=\theta_{\beta}\sigma_{\beta'\beta}^{*}(n), \tag{20}$$

$$\sigma_{\beta\beta'}(n+\alpha N/p)=e^{-2\pi j\alpha(\beta-\beta')/p}\sigma_{\beta\beta'}(n), \tag{21}$$

$$\sigma_{\beta\beta'}(-n)=\sigma_{\beta\beta'}^{*}(n). \tag{22}$$

That the determinant in Eq. (19) is real for all values of k can be shown by demonstrating that it is identically equal to its complex conjugate. The complex conjugate may be formed by replacing each element of the determinant by its conjugate. If this is done and rows and columns are interchanged, then according to (20) one obtains the same determinant as in (19) except that each element $C_{\beta\beta'}$ is multiplied by $\theta_{\beta'}/\theta_{\beta}$. This, however, means that each row is multiplied by $\theta_{\beta'}$ and each column by $1/\theta_{\beta}$ (where β' refers to the row index and β to the column index) which is equivalent to multiplying the entire determinant by $(\theta_1\cdots\theta_{p-1})/(\theta_1\cdots\theta_{p-1})$ or 1. Thus the determinant must be real. It is then apparent that replacing n by $-n$ leaves the determinant unchanged, as according to (22) it merely replaces each element by its complex conjugate. Furthermore, the determinant in (19) is unchanged if n is replaced by $n+\alpha N/p$, where α is any integer. According to (21) this results in multiplying each element $C_{\beta\beta'}$, by

$$e^{-2\pi j\alpha(\beta-\beta')/p}=e^{-2\pi j\alpha\beta/p}e^{2\pi j\alpha\beta'/p}$$

so that each column is multiplied by $e^{-2\pi j\alpha\beta/p}$ and each row by $e^{2\pi j\alpha\beta'/p}$. This is equivalent to multiplying the entire determinant by

$$\prod_{\beta=0}^{p-1} e^{-2\pi j\alpha\beta/p}\prod_{\beta'=0}^{p-1} e^{2\pi j\alpha\beta'/p}$$

which is clearly equal to one.

It is thus immediately apparent that as a result of the properties derived above all possible values are obtained for the determinant in (19) by using for n only the values $0, 1, \cdots, N/2p$ (or $(N-p)/2p$ if $N/2p$ is not an integer). The various roots for each value of n can be designated by n_p for the $p+1$ root, analogous to the symmetric case. In some cases, however, it is customary to regard any asymmetric anode block as having been produced by a perturbation of a symmetric block. In this case the modes of asymmetric block can be thought of as joining continuously with those of the symmetric block, and one retains the designation of the modes of the symmetric block for those of the asymmetric block. Such a designation is unambiguous only if the roots of Eq. (19) corresponding to a given value of n never cross each other while the perturbation is carried out.

2. The Electromagnetic Field and Degeneracies

The electric and magnetic field distributions of the various modes are given by

$$E_{\phi}=\sum_{\beta=0}^{p-1}\frac{N\theta_{\beta}}{\pi p}A_{\beta}\sum_{m=-\infty}^{\infty}\frac{\sin\gamma\theta_{\beta}}{\gamma\theta_{\beta}}\frac{Z_{\gamma}'(k\rho)}{Z_{\gamma}'(ka)}\times e^{-2\pi jm\beta/p}e^{j\gamma\phi},$$

$$E_{\rho}=-(j/k\rho)\sum_{\beta=0}^{p-1}\frac{N\theta_{\beta}}{p\pi}A_{\beta}\times\sum_{m=-\infty}^{\infty}\gamma\frac{\sin\gamma\theta_{\beta}}{\gamma\theta_{\beta}}\frac{Z_{\gamma}(k\rho)}{Z_{\gamma}'(ka)}e^{-2\pi jm\beta/p}e^{j\gamma\phi}, \tag{23}$$

$$H_{z}=-j(\epsilon_0/\mu_0)^{\frac{1}{2}}\sum_{\beta=0}^{p-1}\frac{N\theta_{\beta}}{p\pi}A_{\beta}\times\sum_{m=-\infty}^{\infty}\frac{\sin\gamma\theta_{\beta}}{\gamma\theta_{\beta}}\frac{Z_{\gamma}(k\rho)}{Z_{\gamma}'(ka)}e^{-2\pi jm\beta/p}e^{j\gamma\phi}.$$

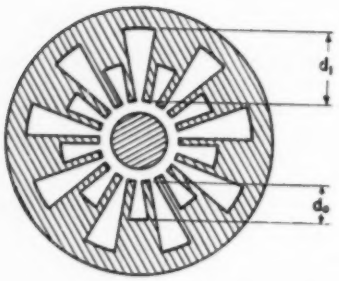


FIG. 8. Vane-type rising sun anode block.

In order to completely determine the above field distributions it is necessary to evaluate the quantities A_β. These were the unknowns in the system of linear homogeneous equations given by (18), and thus only their ratios are determined. These are given by (24) $A_\beta/A_{\beta'}=D_{\beta j}/D_{\beta' j}$, where D_{ij} is the cofactor of the ij'th term of the determinant in Eq. (19).[10]

These co-factors are to be evaluated at the resonant frequency of the mode in question and will thus, in general, be different for different roots of Eq. (19) with a given n as well as for the different values of n.

It is now possible to consider the possibilities for degeneracy. In the previous section it was shown that the set of resonances is unchanged when n is replaced by $-n$ or $n+\alpha(N/p)$ (where α=any integer). One can readily show, however, that replacing n by $n+\alpha(N/p)$ for a particular root of (19) does not give rise to any linearly independent fields, while replacing n by $-n$ gives rise to a linearly independent set of fields unless $n=0$ or $N/2p$.[11] Thus the modes for which $n\neq 0$ or $N/2p$ have a second-order degeneracy which is a consequence of the symmetry of the problem.

In addition to the intrinsic degeneracy pointed out above, accidental degeneracies may appear because of the chance coincidence of a root of Eq. (19) for one value of n with a root for some other value of n, or the appearance of a multiple root in (19) for a particular value of n.[12] The presence of a multiple root is indicated when all of the cofactors in the determinant of Eq. (19) are zero at the resonant frequency in question. Under these conditions the ratios A_β/A_β' are indeterminate, as one would expected for a degeneracy of this type. In general, the multiplicity of a root is equal to the difference between the order and the rank of the matrix whose elements are the same as those of the determinant in Eq. (19). Correspondingly, one finds at such roots that all derivatives with respect to k of this determinant up to the $(m-1)$'th are zero, where m is the order of the multiplicity.

THE RISING SUN TYPE ANODE BLOCK

(1) The Resonant Frequencies

The rising sun type anode block (Fig. 8) is simply the special case in the preceding theory

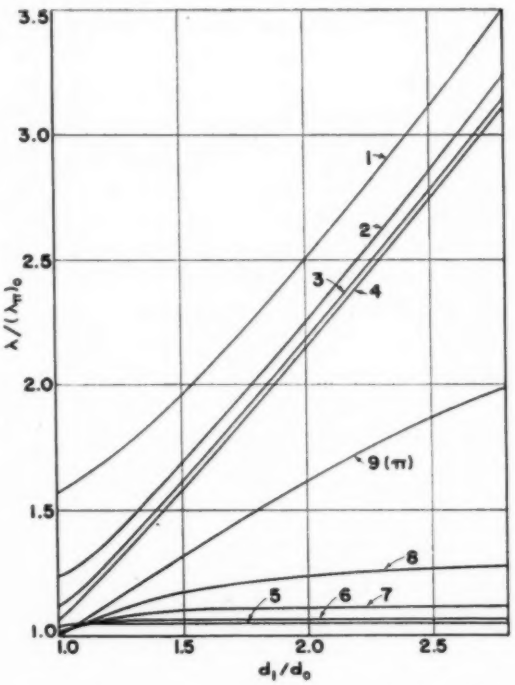


FIG. 9. Spectrum of vane-type rising sun anode block with $N=18$ as a function of d_1/d_0, with d_0 fixed. $(\lambda_\pi)_0$ is the π-mode wave-length for $d_1=d_0$. Dimensions: $2a/(\lambda_\pi)_0=0.524$; $d_0/(\lambda_\pi)_0=0.172$; $c=0$; $\theta=0.060$ radian.

[10] This ratio should of course be independent of j. The requirement that it be independent of j is just equivalent to the requirement that the determinant in (19) be zero.

[11] These results are just what one would expect physically. The mode number, n, determines the phase difference between similar side resonators in successive sections, the phase difference being $(2\pi p/N)n$. Replacing n by $n+\alpha(N/p)$ merely adds $2\pi\alpha$ to this phase difference which, of course, has no physical significance. In addition, when $n=0$ or $N/2p$, replacing n by $-n$ does not result in anything new physically.

[12] These degeneracies are accidental in the sense that they can always be removed by a properly chosen perturbation of the quantities $L_\beta(eq)$.

in which $p=2$. Equation (19) becomes

$$\begin{vmatrix} \sigma_{00}-\cot kL_0 & \sigma_{01} \\ \sigma_{10} & \sigma_{11}-\cot kL_1 \end{vmatrix}=0,$$

or

$$(\sigma_{00}-\cot kL_0)(\sigma_{11}-\cot kL_1)=\sigma_{01}\sigma_{10}, \qquad (25)$$

where

$$\sigma_{00}=\frac{N\theta_0}{2\pi}\sum_{m=-\infty}^{\infty}\left(\frac{\sin\gamma\theta_0}{\gamma\theta_0}\right)^2\frac{Z_\gamma(ka)}{Z_\gamma'(ka)},$$

$$\sigma_{11}=\frac{N\theta_1}{2\pi}\sum_{m=-\infty}^{\infty}\left(\frac{\sin\gamma\theta_1}{\gamma\theta_1}\right)^2\frac{Z_\gamma(ka)}{Z_\gamma'(ka)},$$

$$\theta_1\sigma_{01}=\theta_0\sigma_{10}=\frac{N\theta_1\theta_0}{2\pi}\sum_{m=-\infty}^{\infty}\left(\frac{\sin\gamma\theta_0}{\gamma\theta_0}\right)\times\left(\frac{\sin\gamma\theta_1}{\gamma\theta_1}\right)\frac{Z_\gamma(ka)}{Z_\gamma'(ka)}(-1)^m.$$

As in the case of the symmetric anode this equation is most readily solved graphically. Figure 9 shows how the first two roots for $n\neq 0$, and the second root for $n=0$ (the first being the trivial $k=0$ root) behave as d_0 remains fixed and d_1 is increased starting from $d_1=d_0$, with $N=18$. In this case d_1 and d_0 refer to the vane lengths in the vane-type anode block shown in Fig. 8, so that in Eq. (25) equivalent lengths $L_0(eq)$ and $L_1(eq)$ must be used. It is to be observed that the modes are numbered so as to correspond to the numbering when $d_0=d_1$. That is, the modes 5, 6, 7, 8, and 9 are the second roots for $n=4, 3, 2, 1$, and 0, respectively. Since the first two roots of Eq. (25) for any n do not cross (i.e., the 1 mode and 8 mode do not cross, etc.) there is no ambiguity in assigning mode numbers.[13] The spectrum splits into three parts consisting of a group of modes ($n=1$ to $N/4$ or $(N-2)/4$) whose wave-lengths are determined mainly by the dimensions of the large resonators; a group of modes ($n=N/4$ or $(N+2)/4$ to $(N/2)-1$) whose wave-lengths

[13] For dimensions used in practice double roots of Eq. (25) are impossible except for $n=N/4$. This follows from the fact that σ_{01} and σ_{10} are never zero. For the case $n=N/4$, $\sigma_{01}\equiv\sigma_{10}\equiv 0$, in which case double roots are possible. The notation remains unambiguous, however, because the root is also double for $L_0=L_1$. Thus if $N/4$ is an integer there will be two non-degenerate $N/4$ modes originating from the doublet $N/4$ mode present when $L_1=L_0$.

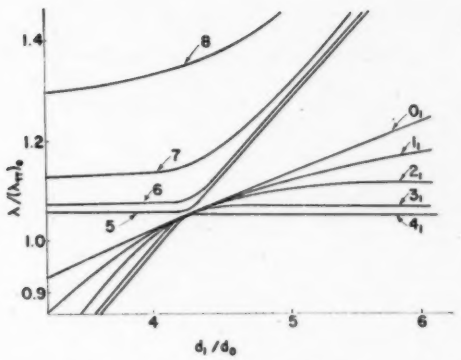


FIG. 10. Behavior of the lower multiplet modes of the rising sun anode block of Fig. 9 for large values of d_1/d_0.

are determined mainly by the dimensions of the small resonator system; and the $N/2$ or π mode which lies between these two groups, and whose wave-length depends more or less equally on the dimensions of both sets of resonators. The longer and shorter wave-length groups are generally referred to as the upper and lower multiplets, respectively. An inspection of Fig. 9 will show that the upper and lower multiplets bear a suggestive resemblance to the first set of resonances of symmetric anode blocks having $N/2$ resonators with dimensions corresponding to the large and small resonators, respectively. A more detailed examination of Eq. (25) shows that this is to be expected. (26) $\sigma_{00}-\cot kL_0=0$ and (27) $\sigma_{11}-\cot kL_1=0$ are the resonance equations for the symmetric $N/2$ resonator anode blocks formed by filling the $\beta=1$ and $\beta=0$ set of resonators with metal, respectively. Equation (25) can be regarded as a coupling equation between the two sets of resonances, with the strength of the coupling depending upon the magnitude of $\sigma_{01}\sigma_{10}$. When this quantity is zero, as it is for $n=N/4$, there is no coupling at all between the two resonant systems, and the roots are given precisely by (26) and (27). For the dimensions used in practice the coupling is quite small for most of the modes; a reasonably accurate spectrum can usually be obtained using (26) and (27) for all the modes, except the $n=N/2$ and the $n=(N/2)-1$ modes.

Figure 9 shows the behavior of the modes for values of d_1/d_0 which are not too large. For larger values of d_1/d_0 the second group of resonances of the large resonator resonant system occurs at a longer wave-length than that of the first group

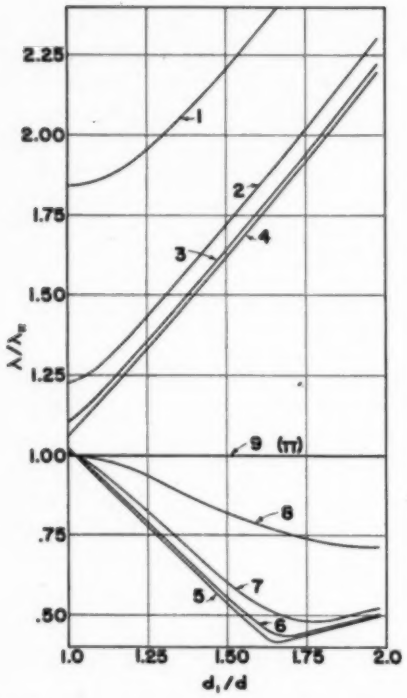


FIG. 11. Spectrum of a vane-type rising sun anode block with $N=18$. Here d_1 and d_0 are simultaneously varied to keep the wave-length of the $n=9$-mode fixed. d is the value of d_1 corresponding to the case $d_1=d_0$. Dimensions: $2a/\lambda_9=0.325$; $d/\lambda_9=0.170$; $c/a=0.595$; $\theta=0.068$ radians.

of resonances of the short wave-length system. The behavior of the 5, 6, 7, and 8 modes at the transition region is shown in Fig. 10.

Inasmuch as the ratio of anode diameter to wave-length of the operating mode is one of the primary parameters in magnetron design, the information given in Fig. 11 is more pertinent to design than that given in Fig. 9. In Fig. 11 the spectrum is plotted as d_1 and d_0 are simultaneously varied in such a way as to keep the π-mode wave-length constant. The abscissa is the ratio d_1/d, where d is the value of d_1 corresponding to the symmetric case $d_1=d_0$. The plot ends when d_0 becomes zero, transition thus being completed from the case of a symmetric anode block of N resonators to that of one with $N/2$ resonators.

(2) The Electromagnetic Fields

For the discussion of the electromagnetic fields it will be assumed that $\theta_0=\theta_1=\theta$. This simplifies the discussion somewhat, and the generalization to the case $\theta_0\neq\theta_1$ can be carried out very easily. With this restriction the electric and magnetic fields are given by

$$E_\phi=\frac{N\theta}{2\pi}\left[(A_0+A_1)\sum_{\substack{m=-\infty\\ m\text{ even}}}^{\infty}\frac{\sin\gamma\theta}{\gamma\theta}\frac{Z_\gamma'(k\rho)}{Z_\gamma'(ka)}e^{j\gamma\phi}+(A_0-A_1)\sum_{\substack{m=-\infty\\ m\text{ odd}}}^{\infty}\frac{\sin\gamma\theta}{\gamma\theta}\frac{Z_\gamma'(k\rho)}{Z_\gamma'(ka)}e^{j\gamma\phi}\right],$$

$$E_\rho=-(j/k\rho)\frac{N\theta}{2\pi}\left[(A_0+A_1)\sum_{\substack{m=-\infty\\ m\text{ even}}}^{\infty}\gamma\frac{\sin\gamma\theta}{\gamma\theta}\frac{Z_\gamma(k\rho)}{Z_\gamma'(ka)}e^{j\gamma\phi}+(A_0-A_1)\sum_{\substack{m=-\infty\\ m\text{ odd}}}^{\infty}\gamma\frac{\sin\gamma\theta}{\gamma\theta}\frac{Z_\gamma(k\rho)}{Z_\gamma'(ka)}e^{j\gamma\phi}\right], \quad (28)$$

$$H_z=-j(\epsilon_0/\mu_0)^{\frac{1}{2}}\frac{N\theta}{2\pi}\left[(A_0+A_1)\sum_{\substack{m=-\infty\\ m\text{ even}}}^{\infty}\frac{\sin\gamma\theta}{\gamma\theta}\frac{Z_\gamma(k\rho)}{Z_\gamma'(ka)}e^{j\gamma\phi}+(A_0-A_1)\sum_{\substack{m=-\infty\\ m\text{ odd}}}^{\infty}\frac{\sin\gamma\theta}{\gamma\theta}\frac{Z_\gamma(k\rho)}{Z_\gamma'(ka)}e^{j\gamma\phi}\right].$$

The character of the fields is evidently dependent upon the relative values of A_0 and A_1. The ratio A_0/A_1 can be found by applying (24) thus,

$$\frac{A_0}{A_1}=-\frac{\sigma_{11}-\cot kL_1}{\sigma_{01}}=-\frac{\sigma_{10}}{\sigma_{00}-\cot kL_0}. \quad (29)$$

When $L_1=L_0$ one has $A_1=\pm A_0$, which, according to (28), gives the r-f fields corresponding to the n and $(N/2)-n$ modes in a symmetric N resonator anode block. For $L_1=0$, $A_1=0$; while for $L_0=0$, $A_0=0$, which is just what one would expect physically. The r-f fields, of course, correspond to those of the n-modes of $N/2$ resonator symmetric anode blocks with resonator lengths L_0 and L_1, respectively. For the values of L_1/L_0 generally used in rising sun magnetrons, the situation is as follows. Consider first the limiting case of small anode diameter (i.e., $(2\pi a/\lambda)<1$). Then for the upper multiplet A_0/A_1 is always small but positive, being largest for $n=1$ and steadily decreasing as n approaches $N/4$ (or $(N-2)/4$). The fact that A_0/A_1 remains positive can be thought of as the vestigial remains of its character when $L_1=L_0$. For the lower multiplet

A_1/A_0 is always small but negative; here the absolute value decreases as n decreases from $(N/2)-1$ to $N/4$ (or $(N+2)/4$). For corresponding modes in the upper and lower multiplet (i.e., modes which are roots of the same frequency equation such as the 1 and $(N/2)-1$ modes) $|A_1/A_0|$ for the lower multiplet modes exceeds A_0/A_1 for the upper multiplet modes. For the π-mode the situation is quite different; A_0/A_1 is near -1 with $|A_0/A_1|$ always less than one. Thus in the case just discussed the π-mode of a symmetric N resonator anode is very little affected by changing the anode into a rising sun anode while the other modes all become like the modes of symmetric $N/2$ vane anodes. If one now increases the anode diameter, A_0/A_1 increases for the upper multiplet modes, $|A_1/A_0|$ increases for the lower multiplet modes, and $|A_0/A_1|$ decreases for the π-mode. Thus the difference in this respect between the behavior of the π-mode fields and that of the fields of the other modes becomes less marked as the anode diameter is increased. Considering now the behavior of A_0/A_1 as the resonator lengths L_1 are changed from L_0 to their final value, there is a steady monotonic change from the values corresponding to symmetric N resonator anodes to the values indicated above. Table II shows the values of the appropriate quantities for the 3J31.[14] One can see that the value of A_1/A_0 is very near -1 for the $n=N/2-1=8$-mode. While this would seem to indicate that the spectrum is very much like that of an 8-mode in a symmetric 18-resonator anode, such a conclusion is not correct. According to Eq. (28) the amplitude of the $\gamma=1$ component is given by

$$\frac{A_0+A_1}{Z_1'(ka)}\frac{\sin\theta}{\theta},$$

and while A_0+A_1 is very small, in this particular case $Z_1'(ka)$ is also very small, so that the amplitude of the $\gamma=1$ component is actually quite large. For ordinary magnetrons this behavior (which can be attributed to the proximity of an interaction space resonance defined by $Z_1'(ka)=0$) can occur only for the $(N/2)-1$ mode, so that for the other modes the ratio, A_1/A_0, is a reliable qualitative criterion for the character of the fields.

[14] A 1.25 cm, 18 resonator rising sun magnetron.

TABLE II. Ratio of voltage across weakly excited resonator to that across strongly excited resonator for the modes of an 18-resonator rising sun anode block.

Upper multiplet and π-mode		Lower multiplet	
n	A_0/A_1	n	A_1/A_0
1	.235	8	−.826
2	.099	7	−.234
3	.043	6	−.095
4	.012	5	−.027
9	−.667		

Inasmuch as the character of the r-f field of the π-mode has important effects upon the electronic efficiency, it is worth further consideration. The fact that $|A_0/A_1|<1$ indicates that the tangential electric field is larger at the large resonator openings than it is at the small resonator openings, which in turn implies the existence of a net e.m.f. surrounding the anode. Figure 12a shows the charge, current, and magnetic field distribution for the π mode of a rising sun N resonator anode block. Whereas the tangential electric fields at the even and odd resonator openings are opposite in direction, the magnetic fields are in the same direction. This would indicate that the magnetic field in the interaction space is unidirectional or that there is a net magnetic flux linking the anode circumference, which varies sinusoidally in time. One can think of the change of this flux as inducing the e.m.f.

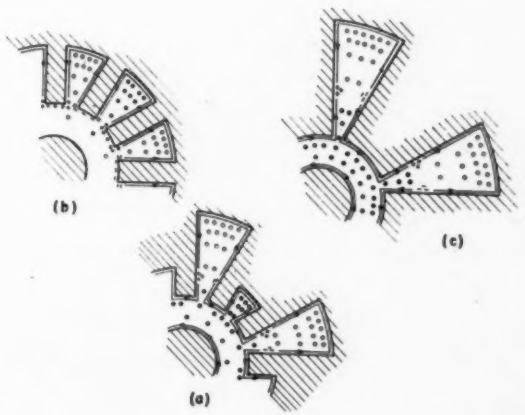


FIG. 12. Distribution of magnetic field, current, and charge for (a) the $N/2$-mode of a N resonator rising sun anode block, (b) the $N/2$-mode of a N resonator symmetric anode block, and (c) the O_1-mode of a $N/2$ resonator symmetric anode block. ○○○ Magnetic field out from paper. ••• Magnetic field into paper.

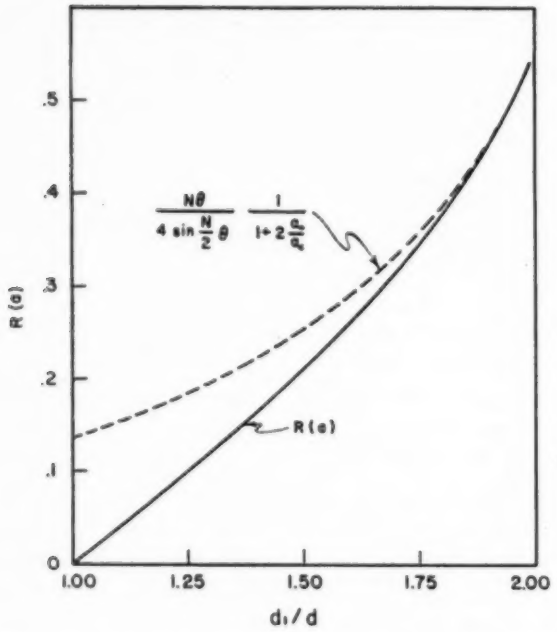


FIG. 13. Zero-mode contamination of the π-mode field corresponding to the spectrum of Fig. 11.

It is also apparent that the current across the various anode segments is unidirectional. Figures 12b and 12c show the charge, current, and magnetic field distribution for the π mode of a symmetric N resonator anode block and for the 0_1 mode of a symmetric $N/2$ resonator anode block. The π-mode of the rising sun anode block bears a certain resemblance to each and can be regarded as a perturbation of either.

As a measure of the zero-mode contamination of the π-mode field it is customary to use the ratio of the $\gamma=0$ component to the $\gamma=N/2$ component of E_ϕ. This ratio is given by

$$R(\rho)=\frac{Z_0'(k\rho)Z_{N/2}'(ka)}{Z_0'(ka)Z_{N/2}'(k\rho)}\frac{A_0+A_1}{A_0-A_1}\frac{N\theta}{4\sin(N/2)\theta}. \quad (30)$$

For rising sun magnetrons in common use this quantity is in the neighborhood of 0.1 at $\rho=a$. As ρ approaches c, $R(\rho)$ increases and may even exceed one in the vicinity of the cathode. Figure 13 shows the behavior of $R(a)$ corresponding to the spectrum of Fig. 11, which shows how the contamination increases as the ratio of resonator lengths is increased. One can readily show that for any ratio of resonator lengths

$$R(a)<\frac{N\theta}{4\sin(N/2)\theta}\frac{1}{1+2(a_0/a_c)} \quad (31)$$

where a_0 is the total area of the small resonators and a_c is the area of the interaction space, and that $R(a)$ approaches this value as the ratio of resonator lengths is increased. This is shown by the dotted curve in Fig. 12. The relation (31) leads one to suspect that a reduction of the area of the interaction space, either by increasing the cathode diameter or decreasing the anode diameter, should reduce the zero-component contamination, and this is, in fact, the case.

From the point of view of magnetron operation the important characteristics of the rising sun resonant system are the separation of the π-mode from the $(N/2)-1$ mode, the separation of the π-mode from the members of the upper multiplet, and the amount of zero-component contamination in the π-mode field. The effects of the various geometric parameters upon these factors can be readily computed using the above theory.[15]

In addition to the above characteristics which refer specifically to the rising sun type anode, there are other electrical properties of general importance in magnetron operation to which the above theory can be applied. Two of these are the unloaded Q and the characteristic admittance (frequently referred to as $(C/L)^{\frac{1}{2}}$). The former has the usual meaning for a cavity resonator, being defined by

$$Q_u=\frac{\omega\ \text{total stored energy}}{\text{average rate of power dissipation in metal walls}}=\omega\kappa\delta\mu_0\frac{\int_V H^2dv}{\int_S H^2ds},$$

where κ is the conductivity and δ the skin depth of the walls. The characteristic admittance is defined by

$$(C/L)^{\frac{1}{2}}=\frac{2\omega(\text{stored energy})}{V^2},$$

where V is the average of the voltage amplitudes across the large and small resonators (i.e.,

[15] A fuller discussion of the effects of the main design parameters will appear in "Microwave magnetrons" of the Massachusetts Institute of Technology Radiation Laboratory Technical Series, Chapter III.

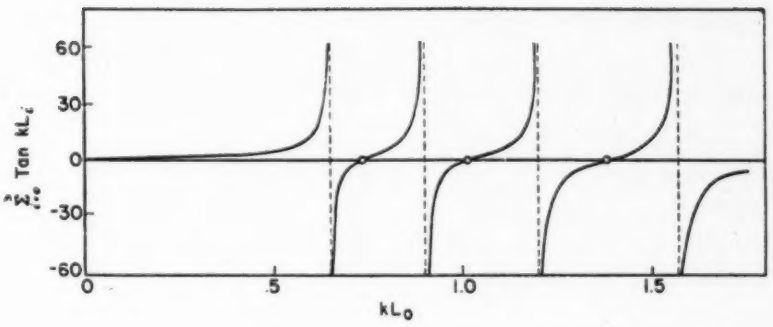


FIG. 14. $\sum_{i=0}^{p-1} \tan kL_i$ plotted as a function of kL_0 for $p=4$, $L_1=1.3L_0$, $L_2=1.75L_0$, $L_3=2.4L_0$.

$V=a\theta(|A_0|+|A_1|)$), and is useful as a means of determining the amplitude of the r-f fields in an operating magnetron. Both of these quantities can be computed using the field distributions given by Eq. 28.[16]

APPROXIMATE SOLUTIONS FOR THE GENERAL CASE

The simple interpretation of the rising sun spectrum suggests that a simple interpretation may also be possible in the general case. There are certain limiting cases for which an interpretation quite analogous to that for the rising sun spectrum is possible. The most important of these is the case in which the anode radius is small compared to the wave-length (e.g., low voltage magnetrons). In this limit, for $n\neq 0$ the σ functions are of the order of a/λ and can therefore be regarded as small quantities. For $n=0$, the σ functions are of the order λ/a so that $1/\sigma$ can be regarded as a small quantity, and, in addition, $\sigma_{ij}/\theta_i-\sigma_{kl}/\theta_k$ is a quantity of the order of a/λ and is thus of the same order of magnitude as $1/\sigma$. Another limiting case of less interest is that of small θ. For this case the σ functions are small for all values of n (of the order of $\theta\ln\theta$). One can use these facts to simplify the equation which determines the resonant frequencies (Eq. 19).

The case for which the σ functions are small will be considered first. If all of the L_i are different one can, as an approximation, ignore all terms involving products of three or more of the σ functions. Under these conditions Eq. (19), which determines the resonant frequencies, reduces to

$$\prod_{i=0}^{p-1}\cot kL_i-\sum_{j=0}^{p-1}\sigma_{jj}\prod_{\substack{i=0\\ i\neq j}}^{p-1}{}'\cot kL_i+\sum_{\substack{k,j=0\\ k<j}}^{p-1}\sigma_{kk}\sigma_{jj}\prod_{\substack{i=0\\ i\neq j\\ i\neq k}}^{p-1}{}'\cot kL_i-\sum_{\substack{k,j=0\\ k<j}}^{p-1}\sigma_{jk}\sigma_{kj}\prod_{\substack{i=0\\ i\neq j,k}}^{p-1}{}'\cot kL_i=0. \quad (32)$$

At any root of (32) one of the $\cot kL_i$, say $\cot kL_\beta$, must be of the same order of magnitude as the σ functions, and, therefore, terms involving products of two σ's with $\cot kL_\beta$ should be omitted for this particular root. Making use of this fact, dividing (32) by

$$\prod_{\substack{i=0\\ i\neq\beta}}^{p-1}{}'\cot kL_i,$$

and rearranging terms, (32) becomes

$$\cot kL_\beta=\sigma_{\beta\beta}-\sum_{\substack{j=0\\ j\neq\beta}}^{p-1}{}'\frac{\sigma_{jj}(\sigma_{\beta\beta}-\cot kL_\beta)}{\cot kL_j}+\sum_{\substack{j=0\\ j\neq\beta}}^{p-1}{}'\frac{\sigma_{j\beta}\sigma_{\beta j}}{\cot kL_j}. \quad (33)$$

It is evident that at the roots of (33), $\cot kL_\beta$ differs from $\sigma_{\beta\beta}$ by terms of the second order in the σ's. Therefore the term

$$\sum_{\substack{j=0\\ \neq\beta}}^{p-1}{}'\frac{\sigma_{jj}(\sigma_{\beta\beta}-\cot kL_\beta)}{\cot kL_j}$$

is of the third order in the σ's and should be

[16] Explicit formulae, derived as indicated above, will appear in "Microwave magnetrons," Chapter III.

omitted. Thus one obtains

$$\cot kL_\beta = \sigma_{\beta\beta} + \sum_{\substack{j=0 \\ j\neq\beta}}^{p-1} \frac{\sigma_{j\beta}\sigma_{\beta j}}{\cot kL_j} \quad (\beta = 0, 1, \cdots, p-1) \quad (34)$$

as the approximate equation for the roots. Equation (34) has a very simple physical interpretation. To the first order in σ, it reads

$$\cot kL_\beta = \sigma_{\beta\beta}, \quad (34')$$

which is the equation for the resonant frequencies of the symmetric anode block formed by filling all resonators, except those designated by L_β, with metal. The term quadratic in the σ's represents the perturbing effect of the other resonators on the roots of (34'). The range of validity of this approximation depends upon the values of the $\cot kL_j$ in the vicinity of the roots. If some of these are of the same order of magnitude as the σ's, then Eq. (34) no longer holds. However, as long as all of the L_i differ, there is always a range of values for the σ's (e.g., for sufficiently small θ_i, or a/λ) for which (34) holds, the range becoming smaller, the nearer two of the L_i approach each other.

In case two or more of the L_i are equal one must include terms of order higher than the second in the σ's. Suppose, for example, that $L_\beta = L_\gamma = L_\delta$. Since there will be one root for which $\cot kL_\beta$ is of the same order of magnitude as the σ's, all of the terms in (32) will, at this root, be of the order of magnitude of terms of cubic or higher order in the σ's. There are many terms in (19) of this order of magnitude which do not appear in (32), so that (32) cannot be correct even as a limiting case. The proper procedure is the following one. Suppose r of the L_i are equal, and L_β is one of these. Then when considering the root for which $\cot kL_\beta$ is of the order of magnitude of the σ's, it is necessary to include all terms of Eq. (19) containing products of $r+1$ or less σ functions, if an approximation to the same order as appears in Eq. (34) is desired. The terms of the order r or less give rise to precisely the resonance equation that one would obtain if all of the other L_i (i.e., the $L_i \neq L_\beta$) were set equal to zero, while the terms of order $r+1$ give the perturbation resulting from the presence of these other resonators.[17] As an example of the above, consider the case in which two of the L_i, say L_β and L_γ, are equal. The approximate solution in this case is

$$\begin{vmatrix} \sigma_{\beta\beta} - \cot kL_\beta & \sigma_{\gamma\beta} \\ \sigma_{\beta\gamma} & \sigma_{\gamma\gamma} - \cot kL_\gamma \end{vmatrix} = \sum_{\substack{j=0 \\ j\neq\beta,\gamma}}^{p-1} \frac{\sigma_{j\gamma}\sigma_{\gamma j}\cot kL_\beta + \sigma_{j\beta}\sigma_{\beta j}\cot kL_\gamma + \sigma_{j\beta}\sigma_{\beta\gamma}\sigma_{\gamma j} + \sigma_{j\gamma}\sigma_{\gamma\beta}\sigma_{\beta j} - \sigma_{\beta j}\sigma_{j\beta}\sigma_{\gamma\gamma} - \sigma_{\gamma j}\sigma_{j\gamma}\sigma_{\beta\beta}}{\cot kL_j}, \quad (35)$$

while in the case all $L_i = 0$ except L_γ and L_β the resonance equation is

$$\begin{vmatrix} \sigma_{\beta\beta} - \cot kL_\beta & \sigma_{\gamma\beta} \\ \sigma_{\beta\gamma} & \sigma_{\gamma\gamma} - \cot kL_\gamma \end{vmatrix} = 0.$$

In case $\theta_\gamma = \theta_\beta$, then $\sigma_{\beta\beta} = \sigma_{\gamma\gamma}$, and (35) can be reduced to a form similar to (34), namely,

$$\cot kL_\beta = \sigma_{\beta\beta} \pm |\sigma_{\gamma\beta}| \pm \frac{1}{2|\sigma_{\gamma\beta}|} \cdot \times \sum_{\substack{j=0 \\ j\neq\gamma,\beta}}^{p-1} \frac{\sigma_{j\beta}\sigma_{\beta\gamma}\sigma_{\gamma j} + \sigma_{j\gamma}\sigma_{\gamma\beta}\sigma_{\beta j} \pm |\sigma_{\gamma\beta}|(\sigma_{\gamma j}\sigma_{j\gamma} + \sigma_{\beta j}\sigma_{j\beta})}{\cot kL_j},$$

where the upper set of signs gives one set of roots and the lower set of signs another set of roots.

The case of σ large is somewhat more complicated. For all L_i different Eq. (19) can be reduced to[18]

$$\sum_{i=0}^{p-1} \theta_i \tan kL_i = \frac{\theta_0}{\sigma_{00}} + \frac{\theta_0}{\sigma_{00}} \sum_{\substack{k,j=0 \\ k<j}}^{p-1} \begin{vmatrix} \sigma_{kk} & \sigma_{kj} \\ \sigma_{jk} & \sigma_{jj} \end{vmatrix} \times \tan kL_j \tan kL_k. \quad (36)$$

Both terms on the right are of the order of $1/\sigma$, since $\sigma_{kk} - \sigma_{kj}$ and $\sigma_{jk} - \sigma_{jj}$ are of the order of $1/\sigma$. The situation here is quite different from that

[17] See Appendix III.
[18] See Appendix III.

which occurs for small σ since the resonators of different lengths are strongly coupled together. Figure 14 shows a qualitative plot of

$$\sum_{i=0}^{p-1} \theta_i \tan kL_i$$

for $p=4$. There will be three resonances corresponding to the first group of modes, and since the right side of Eq. (36) is small (except near the asymptote) these will occur near the three indicated zeros of

$$\sum_{i=0}^{p-1} \theta_i \tan kL_i.$$

In case two of the L_i are nearly equal (say L_0 and L_1) then the root lying between the asymptote of L_0 and L_1 will occur near an asymptote, and the right side of Eq. (36) will no longer be small. This suggests that a different expression will be needed for certain of the roots, in the case that some of the L_i are equal. However, those roots of (36) for which the right side is small will still be correctly given by (36).

Now suppose that r of the L_i are equal to some length L. The determinant in Eq. (19) is a polynomial of the r'th degree which can, therefore, be written as a product of r factors each containing $\cot kL$ linearly. Thus Eq. (19) can be replaced by r equations involving $\cot kL$ linearly. $r-1$ of these will have sets of roots for which $\cot kL$ is of the order of $1/\sigma$, and to the first order in $1/\sigma$ these roots are the same as they would be if all of the other L_i ($L_i \neq L$) were set equal to zero.[19] The effects of the other resonators appear in the higher order terms. These roots thus behave in the same way as those for which σ is a small quantity.

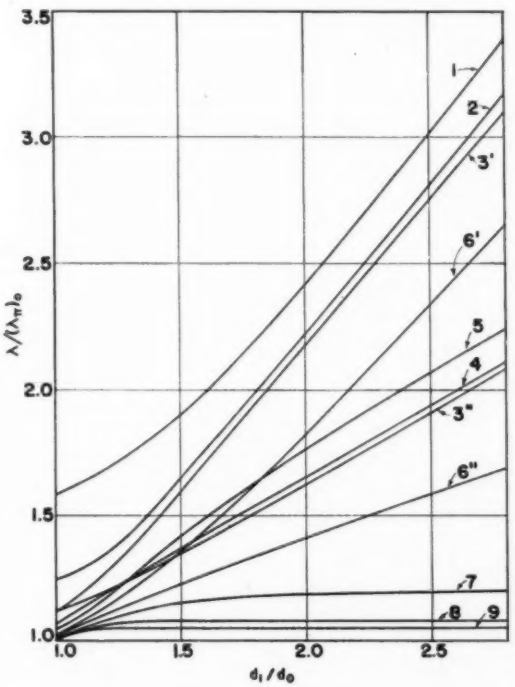


FIG. 15. Spectrum of magnetron anode block with $p=3$, $N=18$ as a function of L_2/L_0. Here $L_1=\frac{1}{2}[(L_2/L_0)+1]$ and L_0 is held fixed.

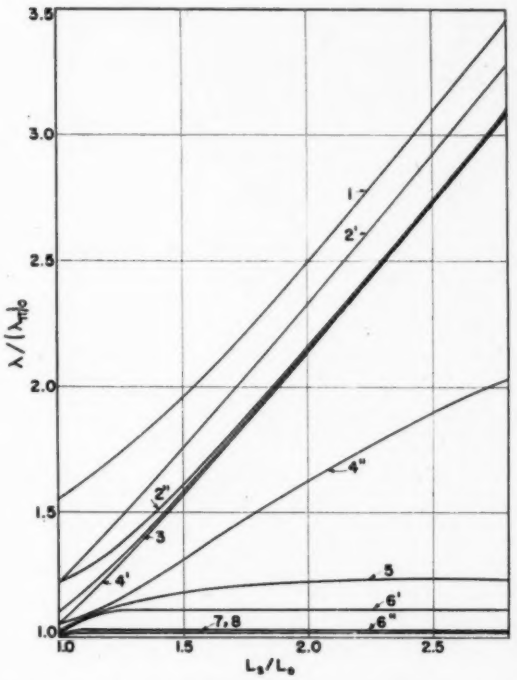


FIG. 16. Spectrum of magnetron anode block with $p=4$, $N=16$ as a function of L_2/L_0. Here $L_1=L_0$, $L_3=L_2$, and L_0 is held fixed.

Two illustrative examples of the preceding discussion are shown in Figs. 15 and 16. Figure 15 shows qualitatively the behavior of the spectrum of an anode with $p=3$, $N=18$, and $L_1/L_0=\frac{1}{2}[(L_2/L_0)+1]$ given as L_2/L_0 varies from 1 to 3 with L_0 fixed. There are three groups of modes corresponding to the spectrum of the six resonator symmetric anodes with resonator lengths L_0, L_1, and L_2, respectively. Any one of these groups is left practically unchanged both as regards to frequencies and field patterns if the

[19] See Appendix III.

TABLE III.* Comparison between measured and computed wave-lengths of a closed-end, vane-type, rising sun anode block; block dimensions: $a=2.296$ cm; $d_0/a=.6571$; $d_1/a=1.208$; $h/a=3.267$; $\theta=.060$ radians.

Mode number	Computed wave-lengths (cm)	Measured wave-lengths (cm)
1	12.07	12.07
2	11.55	11.60
3	11.39	11.42
4	11.32	11.37
9	9.96	10.05
8	8.72	8.72
7	8.14	8.11
6	7.91	7.88
5	7.85	7.80

* The authors are indebted to Mr. Edwin Richter for making the above measurements.

two sets of resonators not involved are omitted entirely.[20] In addition there are the two $n=6$ modes, well separated from the other modes and involving all of the resonators. These modes are analogous to the π mode of a rising sun magnetron. It should be observed that the qualitative aspects of the spectrum, such as order of modes and location of large separations, can be predicted with no calculation at all, while a rough quantitative estimate of mode separation can be made with very little calculation.

For the anode of Fig. 16, $N=16$, $p=4$, $L_0=L_1$, and $L_2=L_3$, the spectrum is given as a function of L_2/L_0, with L_0 fixed. Here one has two groups of modes, one associated with the large set of resonators, the other with the small set of resonators, and a well separated mode in between the two groups. The well separated mode, while analogous in position to the π-mode, has no π ($\gamma=8$) component. This problem is somewhat more complicated than the preceding one in that some calculation is needed to ascertain the order of the modes in the two groups.

One can use the theory presented above to consider quite generally the advantages which can be obtained using various resonators of different lengths for magnetron anodes. First of all, the only modes which involve appreciably more than one size resonators are certain of the roots of Eq. (19) for which $n=0$, there being one such mode lying between each pair of zero's of the $\cot kL_i$. These modes may or may not be "π" modes (π mode here means a mode containing a substantial $\gamma=N/2$ component). In those cases for which one has a well separated π mode, the interaction field of these modes will contain, in general, some zero component and also components with $\gamma=m(N/p)$. Since some of these components will have $\gamma<N/2$, the interaction field near the cathode may have, relatively, much less π component than it has near the anode. Thus it appears that from the point of view of quality of the field any such structure will probably be worse than the rising sun and will certainly be no better. One can reduce the number of long wave-length modes for a given value of N by using arrangements different from the rising sun. However, in such cases one always simultaneously increases the possibilities of harmonic operation in the long wave-length modes so that there may be no over-all improvement. An attempt has been made to cure long wave-length moding problems in this way and has failed.[21]

While one can show that the rising sun is almost certainly the best kind of unstrapped resonant system making use of unequal cavities, it is not apparent that a satisfactory resonant system using all equal cavities, but not evenly spaced, cannot be designed. It has been shown, of course, that the mode separation is small in the limit of low voltage, but even within these limitations it might be possible, though difficult, to obtain sufficient mode separation for satisfactory operation. One will have, however, in this case also, low γ-field components in the interaction field which may seriously impair the efficiency of operation.

END SPACE EFFECTS AND THE ACCURACY OF THE CALCULATION

As noted previously the preceding theory is based upon the assumption that the axial variation of the fields in the anode block can be neglected. The validity of this assumption depends upon the anode height, the end space height, and the end space diameter. For sufficiently large anode height, there are always

[20] This was shown explicitly only with regard to frequency (Eq. 34). That the field patterns are also unchanged follows from the fact that the field intensities in all of the resonators except that for which $\cot kL$ is small are very small, a fact which can be ascertained by examining the cofactors of the determinant in Eq. 19.

[21] N.D.R.C. Report No. 588, Section 2T.

modes for which the assumption is valid.[22] In addition, for modes of large n, the end space (considered as a section of coaxial wave guide) is beyond cutoff so that the effect of the fields of the end space are confined to a region near the ends of the anode block. If the end space height is larger than the penetration of these fields, the effect of the end space may be small even when the anode block itself is not particularly long.[23]

For modes of low n the end space may not be beyond cutoff; in this case the magnitude of the effect depends upon the strength of the lowest γ component in the magnetron field. For the 18-resonator rising sun magnetron (3J31), the wavelength of the one mode is depressed approximately 25 percent, while that of the other modes is depressed 5 percent or less.

In addition to the end space approximation the theory contains a further inaccuracy because of the inexact matching at the junctions between the resonator openings and the interaction space. The error introduced thereby can be measured quite accurately if the end spaces can be chosen so as to admit an exact solution. The latter can be accomplished by entirely covering both ends of the anode block with metal. The magnetron anode block, being cylindrical, can be thought of as a section of wave guide, with cut-off wavelengths given by the open-end wave-lengths previously computed. The resonant wave-lengths (λ_{nc}) of the closed-end magnetron are thus given by the standard formula for the resonances of a half wave-length cylindrical cavity, namely,

$$\lambda_{nc}=\frac{\lambda_n}{(1+(\lambda_n/2h)^2)^{\frac{1}{2}}},$$

where λ_n is the cut-off wave-length for the mode in question and h is the anode height. An experiment of this kind has been performed using a vane-type rising sun anode block. A comparison of the observed and computed wavelengths appears in Table III. It is to be observed that the computed wave-lengths of the 1, 2, 3, 4, and 9 modes tend to be short, while those of the 5, 6, 7, and 8 tend to be long. This behavior has been traced to a machining error in the construction of the anode. Rough calculations indicate that a substantial part of the differences which appear in Table III may be due to this machining error. The matter has not been pursued further because the experiment clearly indicates that the matching error is quite negligeable compared to end space effects. For the same reason no extensive effort has been made to estimate the error mathematically or to improve the matching (say by using a better approximation to E_ϕ at the resonator openings).

[22] For this reason this type of calculation is often referred to as the infinite height approximation, a name which is somewhat misleading since an infinitely long anode would not have discrete resonant frequencies.

[23] Experimental results on dependence of the spectrum on the end space height and anode height can be found in Rad. Lab. Report 43-9, "Resonant modes of the magnetron," by J. C. Slater.

APPENDIX I

Some justification for the matching procedure used in this paper can be obtained from the variational principal established by the following theorem.

Consider a closed cavity of arbitrary geometry with conducting walls and divided into two regions, I and II, by an arbitrary surface, S. One may now seek in region I and II, separately, a solution to the vector wave equation $\nabla^2\mathbf{A}+k^2\mathbf{A}=0$ satisfying the conditions that (1) $\mathbf{A}$ and its derivative be continuous throughout the regions I and II but not necessarily across the dividing surface S. (2) $\nabla\cdot\mathbf{A}=0$ and (3) $\mathbf{A}\times\mathbf{n}=0$ along the metal surfaces where $\mathbf{n}$ is a unit vector normal to the metal surfaces and (4) $\mathbf{A}\times\mathbf{n}_s=\mathbf{V}$ along S where $\mathbf{n}_s$ is a unit vector normal to S directed into region I and the function $\mathbf{V}$ can for the moment be considered to be arbitrary. As long as $\mathbf{V}$ is not identically zero, a unique solution to this problem will, in general exist for any value of k. If, however, one further requires that

$$\int_S[(\mathbf{A}\times\text{curl}\mathbf{A})_I-(\mathbf{A}\times\text{curl}\mathbf{A})_{II}]\cdot\mathbf{n}_s ds=0, \qquad (37)$$

a discrete set of k's will be determined.

Now let $\mathbf{A}_0$ be the vector potential of a resonant mode of the cavity with $k=k_0$. $\mathbf{A}_0$ will satisfy the wave equation, Eq. (37), and conditions (1), (2), and (3) above. Along S, $\mathbf{A}_0\times\mathbf{n}_s=\mathbf{V}_0$. Now take $\mathbf{V}=\mathbf{V}_0+a\mathbf{U}$, where $\mathbf{U}$ is an arbitrary vector function and a an arbitrary parameter. This is equivalent to regarding $\mathbf{V}$ as an arbitrary modification of $\mathbf{V}_0$. For a given $\mathbf{U}$ the k's determined by Eq. (37) can be regarded as functions of a. One of these, which shall be designated as k_0', reduces to k_0 as a approaches zero. It shall now be shown that $(dk_0'/da)_{a=0}=0$.

Proof: Let

$$\mathbf{A}_1=\mathbf{A}-\mathbf{A}_0.$$

Then

$$\nabla^2\mathbf{A}=\nabla^2\mathbf{A}_0+\nabla^2\mathbf{A}_1,$$

which reduces to

$$k^2\mathbf{A}=k_0^2\mathbf{A}_0-\nabla^2\mathbf{A}_1.$$

Taking the scalar product of both sides by $\mathbf{A}$ and integrating

throughout the volume of the cavity yields

$$k^2\int_V A^2dv = k_0^2\int_V \mathbf{A}\cdot\mathbf{A}_0dv - \int_V \mathbf{A}\cdot\nabla^2\mathbf{A}_1dv. \quad (38)$$

Applying the vector identity $\mathbf{a}\cdot\text{curl curl }\mathbf{b}-\mathbf{b}\cdot\text{curl curl }\mathbf{a}=\text{div}[\mathbf{b}\times\text{curl }\mathbf{a}-\mathbf{a}\times\text{curl }\mathbf{b}]$ one finds that

$$\int_V \mathbf{A}\cdot\nabla^2\mathbf{A}_1dv = -k^2\int\mathbf{A}_1\cdot\mathbf{A}dv + \int_V \text{div}[\mathbf{A}\times\text{curl }\mathbf{A}_1-\mathbf{A}_1\times\text{curl }\mathbf{A}]dv,$$

so that Eq. (38) becomes

$$k^2 = k_0^2 - \frac{\int_V \text{div}[\mathbf{A}\times\text{curl }\mathbf{A}_1-\mathbf{A}_1\times\text{curl }\mathbf{A}]dv}{\int_V \mathbf{A}\cdot\mathbf{A}_0dv}. \quad (39)$$

Now the volume integral over the divergence of a vector can be transformed into a surface integral. The regions I and II must, however, be considered separately because of the presence of possible discontinuities at S. Thus

$$\int_V \text{div.}[\mathbf{A}\times\text{curl }\mathbf{A}_1-\mathbf{A}_1\times\text{curl }\mathbf{A}]dv = \int_S[(\mathbf{A}\times\text{curl }\mathbf{A}_1)_{II}-(\mathbf{A}\times\text{curl }\mathbf{A}_1)_I]\cdot\mathbf{n}_s ds - \int_S[(\mathbf{A}_1\times\text{curl }\mathbf{A})_{II}-(\mathbf{A}_1\times\text{curl }\mathbf{A})_I]\cdot\mathbf{n}_s ds.$$

The metal walls do not contribute to the surface integrals since both $\mathbf{A}\times\mathbf{n}$ and $\mathbf{A}_1\times\mathbf{n}$ are zero along these walls. Now

$$\int_S[(\mathbf{A}\times\text{curl }\mathbf{A}_1)_{II}-(\mathbf{A}\times\text{curl }\mathbf{A}_1)_I]\cdot\mathbf{n}_s ds = \int_S[(\mathbf{A}\times\text{curl }\mathbf{A})_{II}-(\mathbf{A}\times\text{curl }\mathbf{A})_I]\cdot\mathbf{n}_s ds - \int_S[(\mathbf{A})_{II}-(\mathbf{A})_I]\times\text{curl }\mathbf{A}_0\cdot\mathbf{n}_s ds = 0,$$

since $\mathbf{A}$ satisfies Eq. (37) and $\mathbf{A}\times\mathbf{n}_s$ has the same value on both sides of the surface S. Furthermore,

$$\int_S[(\mathbf{A}_1\times\text{curl }\mathbf{A})_{II}-(\mathbf{A}_1\times\text{curl }\mathbf{A})_I]\cdot\mathbf{n}_s ds = \int_S[(\mathbf{A}_1\times\text{curl }\mathbf{A}_1)_{II}-(\mathbf{A}_1\times\text{curl }\mathbf{A}_1)_I]\cdot\mathbf{n}_s ds + \int_S[(\mathbf{A}_1)_{II}-(\mathbf{A}_1)_I]\times\text{curl }\mathbf{A}_0\cdot\mathbf{n}_s ds - a\int_S\mathbf{U}\cdot[(\text{curl }\mathbf{A}_1)_{II}-(\text{curl }\mathbf{A}_1)_I]ds,$$

since $\mathbf{A}_1\times\mathbf{n}_s$ has the same value on both sides of S, namely, $a\mathbf{U}$. Thus Eq. (39) becomes

$$k^2 = k_0^2 - a\frac{\int_S\mathbf{U}\cdot[(\text{curl }\mathbf{A}_1)_{II}-(\text{curl }\mathbf{A}_1)_I]ds}{\int_V\mathbf{A}\cdot\mathbf{A}_0dv}.$$

If k is that root of Eq. (37) which approaches k_0 as a approaches zero, (i.e., $k=k_0'$) then $\mathbf{A}_1$ also approaches zero and $\mathbf{A}$ approaches $\mathbf{A}_0$. It is thus apparent that

$$(dk_0'/da)_{a=0}=0.$$

Since the electric field is proportional to $\mathbf{A}$ and the magnetic field to curl $\mathbf{A}$, Eq. (37) is equivalent to the requirement that $\int_S\mathbf{E}\times\mathbf{H}\cdot\mathbf{n}_s ds$ have the same value on both sides of S.[24] If $\mathbf{E}\times\mathbf{n}_s$ is chosen constant along S, this requirement is equivalent to the condition that the average values of $\mathbf{H}\times\mathbf{n}_s$ match at S.

APPENDIX II

The following theorem corroborates the statement that no solutions are lost by assuming that fields in successive similar sections differ only by a constant phase factor.

Suppose a cavity has an axis of symmetry such that a rotation of the cavity through an angle $2\pi\alpha p/N$ about this axis leaves the cavity physically unchanged, where α is any integer and N/p[25] is an integer which characterizes the symmetry. Then any mode of this cavity has the property that either: (1) Each of its field components, say $f(\rho, z, \phi)$, where ρ, z, and ϕ are cylindrical coordinates about the axis of symmetry, has the property that

$$f(\rho, z, \phi) = e^{i\delta\alpha}f\left(\rho, z, \phi-\frac{2\pi\alpha p}{N}\right),$$

where δ is a constant, or (2) the mode is a linear combination of modes having the property (1) and the same resonant frequency as the mode in question.

Proof: It is simplest to express the field in terms of a vector potential $\mathbf{a}$ such that $B=\text{curl }\mathbf{a}$ and $\mathbf{E}=-\dot{\mathbf{a}}=-j\omega\mathbf{a}$. Suppose the cavity has a resonant mode of frequency, ω_0, and field distribution given by $\mathbf{a}_0(\rho, z, \phi)$. Let

$$\mathbf{a}_\alpha(\rho, z, \phi) = \mathbf{a}_0\left(\rho, z, \phi-\frac{2\pi\alpha p}{N}\right).$$

Because of the symmetry, the cavity must also have modes of frequency, ω_0, having the field distribution given by $\mathbf{a}_\alpha$. Furthermore, any linear combination of the fields, $\mathbf{a}_\alpha$, gives a mode of oscillation having the resonant frequency ω_0. Now the linear combinations

$$\mathbf{A}_n(\rho, z, \phi) = \sum_{\alpha=0}^{(N/p)-1}\mathbf{a}_\alpha e^{2\pi j\alpha np/N}, \quad n=0, 1, \cdots, \frac{N}{p}-1$$

have the property

$$\mathbf{A}_n(\rho, z, \phi) = e^{i2\pi jnp/N)\alpha}\mathbf{A}_n\left(\rho, z, \phi-\frac{2\pi\alpha p}{N}\right),$$

as can be seen by inspection. It follows, of course, that the electric and magnetic fields derivable from $\mathbf{A}_n$ also have this property. It follows immediately from the definition of $\mathbf{A}_n$ and the fact that

$$\sum_{n=0}^{(N/p)-1} e^{2\pi jn\alpha p/N} = N/p \text{ for } \alpha=0 = 0 \quad \text{for } \alpha\neq 0$$

that

$$\mathbf{a}_0(\rho, z, \phi) = \frac{p}{N}\sum_{n=0}^{(N/p)-1}\mathbf{A}_n,$$

which proves the theorem.

---

[24] One can readily show from this that Eq. (37) is equivalent to requiring that the energy stored in the magnetic field equal that stored in the electric field.

[25] The integer N/p is written in this form to show the relationship to the cavities discussed in this paper.

APPENDIX III

Certain statements made in the text with regard to approximate solutions of Eq. 19 are proved below.

1. *The case of σ small, and several of the L_i equal:* To simplify the notation the discussion will be restricted to the case $L_0=L_1=\cdots L_{r-1}=L$. The generalization is trivial. First it will be shown that in the case

$$L_r=L_{r+1}=\cdots L_{p-1}=0,$$

(19) reduces to

$$\begin{vmatrix} \sigma_{00}-\cot kL & \sigma_{01} & \cdots & \sigma_{0,r-1} \\ \sigma_{10} & \sigma_{11}-\cot kL & \cdots & \sigma_{1,r-1} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \sigma_{r-1,0} & \sigma_{r-1,1} & \cdots & \sigma_{r-1,r-1}-\cot kL \end{vmatrix} =0. \quad (40)$$

If in (19) the r'th column is divided by $\cot kL_r$, the $r+1$ by $\cot kL_{r+1}$, and so forth up to the $(p-1)$th (this is equivalent to dividing the entire determinant by $\prod_{i=r}^{p-1} \cot kL_i$), and then the L_i $(i=r\cdots p-1)$ go to zero, (19) becomes

$$\left|\begin{array}{ccc|ccc} \sigma_{00}-\cot kL & \cdots & \sigma_{0,r-1} & 0 & \cdots & 0 \\ \cdot & & \cdot & \cdot & & \cdot \\ \cdot & & \cdot & \cdot & & \cdot \\ \sigma_{r-1,0} & \cdots & \sigma_{r-1,r-1}-\cot kL & 0 & \cdots & 0 \\ \hline \sigma_{r,0} & \cdots & \sigma_{r,r-1} & -1 & \cdots & 0 \\ \cdot & & \cdot & \cdot & & \cdot \\ \cdot & & \cdot & \cdot & & \cdot \\ \sigma_{p-1,0} & \cdots & \sigma_{p-1,r-1} & 0 & \cdots & -1 \end{array}\right| =0,$$

which is equivalent to Eq. (40). Equation (40) can be expanded as a polynomial in $\cot kL$ as follows:

$$(-\cot kL)^r+(-\cot kL)^{r-1}\sum_{j=0}^{r-1}\sigma_{jj}+\cdots +(-\cot kL)^{r-s}\sum_{\substack{j_1,j_2,\cdots,j_s \\ j_1<j_2<\cdots<j_s}}\Delta^{(s)}{}_{j_1,\ldots,j_s}+\cdots +\Delta^{(r)}{}_{0,1,\ldots,r-1}=0, \quad (41)$$

where $\Delta^{(r)}{}_{0,1,\ldots,r-1}$ is the r'th ordered determinant

$$\begin{vmatrix} \sigma_{00} & \sigma_{01} & \cdots & \sigma_{0,r-1} \\ \sigma_{10} & \sigma_{11} & \cdots & \sigma_{1,r-1} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \sigma_{r-1,0} & \cdots & \cdots & \sigma_{r-1,r-1} \end{vmatrix}$$

and $\Delta^{(s)}{}_{j_1,\ldots,j_s}$ is the s'th ordered determinant formed by striking out of the determinant $\Delta^{(r)}{}_{0,1,\ldots,r-1}$ all rows and columns whose indices do not appear among the subscripts of $\Delta^{(s)}{}_{j_1,\ldots,j_s}$. Each term of a $\Delta^{(s)}$ consists of a product of s σ's and therefore is a small quantity of order $(\sigma)^s$. The determinant $\Delta^{(s)}$ will, in general, be of order $(\sigma)^s$ also, although it is conceivable that for some special geometry, $\Delta^{(s)}$ may be of higher order in σ. It is clear that if $\Delta^{(r)}$ is of the order of $(\sigma)^r$, all roots of (41) must occur at values of k such that $\cot kL$ is of the same order of magnitude as the σ's. If $\Delta^{(r)}$ is of the order $r+1$ or higher in σ, than (41) may have roots for which $\cot kL$ is of the second or higher order in σ.

For the case in which the $L_r, \cdots, L_{p-1}$ are not equal to zero and also not equal to L, Eq. (19) will still have roots for which $\cot kL$ is of the order of magnitude of the σ's. At such a root Eq. (19) will have terms of order $r, r+1, \cdots$ in the σ's but no terms of order less than r. Each term of Eq. (19) consists of a product of p factors. The only possible factors which are not of the order σ are the $\cot kL_i$ $(i=r, r+1, \cdots, p-1)$ and therefore the terms of order σ^r must all have

$$\prod_{i=r}^{p-1}\cot kL_i$$

as a factor. It is immediately apparent that the totality of terms having

$$\prod_{i=r}^{p-1}\cot kL_i$$

as a factor is given by the determinant in Eq. (40) multiplied by this factor, and, therefore, if the terms of order higher than σ^r are omitted, Eq. (19) becomes the same as Eq. (40). Thus the effect of the resonators r to $p-1$ on the roots for which $\cot kL$ is of the order of σ appears only in terms of order $r+1$ and higher in σ. One can readily show that this means that the value of $\cot kL$ at a root depends upon the L_i $(i=r, \cdots, p-1)$ only in terms of the second or higher order in σ.

The case of a root occurring with $\cot kL$ of the second or higher order in σ can be regarded as a rare accident and is of no particular importance. The effect of the other resonators on the value of $\cot kL$ at resonance will still be of the second order in σ.

2. *The case of σ large and all of the L_i different:*

In addition to assuming the σ large it is necessary to assume that $\sigma_{ij}/\theta_i-\sigma_{kl}/\theta_k$ is of the order of magnitude of $1/\sigma$. As pointed out in the text, for $n=0$ and a/λ small, such is the case.

To find the roots of (19) it is convenient to write it as a polynomial in the $\cot kL_i$, which gives

$$(-1)^p\prod_{i=0}^{p-1}\cot kL_i+(-1)^{p-1}\sum_{j=0}^{p-1}\sigma_{jj}\prod_{\substack{i=0 \\ i\neq j}}^{p-1}\cot kL_i +(-1)^{p-2}\sum_{\substack{k,j=0 \\ k<j}}^{p-1}\begin{vmatrix}\sigma_{kk} & \sigma_{jk} \\ \sigma_{kj} & \sigma_{jj}\end{vmatrix}\prod_{\substack{i=0 \\ i\neq j,\,k}}^{p-1}\cot kL_i+\cdots +(-1)^{p-s}\sum_{\substack{j_1,j_2,\cdots,j_s \\ j_1<j_2<\cdots<j_s}}\Delta^{(s)}{}_{j_1,\ldots,j_s}\prod_{\substack{i=0 \\ i\neq j_1,\cdots,j_s}}^{p-1}\cot kL_i+\cdots +\Delta^{(p)}{}_{0,1,\ldots,p-1}=0, \quad (42)$$

where the Δ's are defined as before, but with respect to the determinant

$$\Delta^{(p)}{}_{0,1,\ldots,p-1}=\begin{vmatrix} \sigma_{00} & \sigma_{01} & \cdots & \sigma_{0,p-1} \\ \sigma_{10} & \sigma_{11} & \cdots & \sigma_{1,p-1} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \sigma_{p-1,0} & \sigma_{p-1,1} & \cdots & \sigma_{p-1,p-1} \end{vmatrix}.$$

Now the $\Delta^{(s)}$ are determinants of the form

$$\begin{vmatrix} \sigma_{j_1 j_1} & \cdots & \sigma_{j_1 j_s} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \sigma_{j_s j_1} & \cdots & \sigma_{j_s j_s} \end{vmatrix}$$

$$= \begin{vmatrix} (\sigma_{j_1 j_1} - \sigma_{j_1 j_2}) & \cdots & (\sigma_{j_1 j_{s-1}} - \sigma_{j_1 j_s}) & \sigma_{j_1 j_s} \\ \cdot & & \cdot & \cdot \\ \cdot & & \cdot & \cdot \\ \cdot & & \cdot & \cdot \\ (\sigma_{j_s j_1} - \sigma_{j_s j_2}) & \cdots & (\sigma_{j_s j_{s-1}} - \sigma_{j_s j_s}) & \sigma_{j_s j_s} \end{vmatrix}, \quad (43)$$

since a determinant is unchanged if any row is subtracted from any other row. Since each element in the right side of (43) is of the order of $1/\sigma$ except those of the last row, it is apparent that $\Delta^{(s)}$ is of the $s-2$ or higher order in $1/\sigma$. It is thus apparent that the largest term in (42) is the second one and the next largest are the first and third Retaining only these and dividing through by

$$\frac{\sigma_{00}}{\theta_0} \prod_{i=0}^{p-1} \cot kL_i,$$

one finds

$$\sum_{i=0}^{p-1} \theta_i \tan kL_i = \frac{\theta_0}{\sigma_{00}} + \frac{\theta_0}{\sigma_{00}} \sum_{\substack{k,j=0 \\ k<j}}^{p-1} \begin{vmatrix} \sigma_{kk} & \sigma_{kj} \\ \sigma_{jk} & \sigma_{jj} \end{vmatrix} \tan kL_k \tan kL_j, \quad (44)$$

which is Eq. (36). On the left side of (44) the coefficients $(\sigma_{jj}/\sigma_{00})\theta_0$ have been replaced by θ_j since they differ from θ_j by

$$\left(\frac{\sigma_{00}}{\theta_0} - \frac{\sigma_{jj}}{\theta_j}\right)\frac{\theta_0 \theta_j}{\sigma_{00}},$$

which is of the order $(1/\sigma)^2$.

3. *The case of σ large and several of the L_i equal:* As in 1, the discussion will be restricted to the case $L_0 = L_1 = \cdots = L_{r-1} = L$. The case in which $L_r = L_{r+1} = \cdots = L_{p-1} = 0$ will be considered first. In this case the resonance equation can be written as (41) where the $\Delta^{(s)}$ are of the $s-2$ or higher order in $1/\sigma$. As in the case considered in 1, while $\Delta^{(r)}$ may be of order higher than $r-2$ in $1/\sigma$, such an occurrence can be considered to be a rare accident. The remarks made concerning the similar possibility in 1 will apply here also. Therefore, in the following discussion it will be assumed that $\Delta^{(r)}$ is of the order of $(1/\sigma)^{r-2}$.

Equation (41) can be regarded as a polynomial in $\cot kL$, which means that it can be written in the form

$$\prod_{i=1}^{r} (\cot kL - a_r) = 0.$$

The a_r will be functions of k and, as can readily be shown, will be real. Thus the resonance condition (41) can be rewritten as

$$\cot kL = a_i \quad (i = 1, \cdots, r). \quad (45)$$

One can see that one of the a_i, say a_r, is of the order of σ. For if at a root of (41) $\cot kL$ is of the order of σ, the first two terms become dominant and one finds that to a first approximation

$$\cot kL = \sum_{j=0}^{r-1} \sigma_{jj}.$$

$\sum_{j=0}^{r-1} \sigma_{jj}$ differs from a_r only in terms of order $1/\sigma$ and higher.

It can further be shown that the remaining a_i, say $a_1 \cdots a_{r-1}$, are of order $1/\sigma$. If at a root of (41) $\cot kL$ is of order $1/\sigma$, then only the first term (and perhaps some of the terms between the second and last) becomes negligible, and a polynomial in $\cot kL$ of degree $r-1$ remains. If one assumes any other kind of behavior for $\cot kL$, either the second or the last term becomes dominant and it is not possible to satisfy Eq. (41).

Now consider the case in which the $L_r \cdots L_{p-1}$ can have any value except L. Considering the resonance equation in the form (42), and assuming $\cot kL$ to be of the order of $1/\sigma$, one finds that the dominant terms are all of the order of $(1/\sigma)^{r-2}$ and that they are the same as the dominant terms in (41) except that they are all multiplied by the common factor

$$\prod_{i=r}^{p-1} \cot kL_i.$$

From this one can readily show that the equations in (45) for $i = 1, \cdots, r-1$ become replaced by $\cot kL = a_i + b_i$, where b_i is a quantity of order $(1/\sigma)^2$ and contains the dependence of the frequency upon the resonators L_r to L_{p-1}. Thus these roots behave quite similarly to those of Eq. (19) which one finds for σ small.

Since the only assumption required in the derivation of Eq. (44) is that none of the $\cot kL_i$ approach zero in the neighborhood of a root, Eq. (44) is valid for any roots satisfying this requirement even when some of the L_i are equal. One can readily see that, neglecting terms of order $(1/\sigma)$, Eq. (44) reduces to (45) with $i = r$ for $L_r = L_{r+1} = L_{p-1} = 0$.

The Palletron, A New Electron Resonator and Its Proposed Application to the Generation of Potentials in the Million-Volt Range*

A. M. SKELLETT
National Union Radio Corporation, Orange, New Jersey

(Received September 2, 1947)

The kind of electrostatic field in which an electron or ion will follow simple harmonic motion is shown to have a parabolic potential distribution. A gap at the center of this field provides a means for exchange of energy between the electrons and the associated circuit. If energy is taken from the electrons the device provides an oscillator of high efficiency; if energy is given to them the device may be used to generate a high d.c. voltage. Both kinds of device have been operated successfully.

Experimental results on a small model of the voltage generator are given and a proposed design for the million-volt range is briefly described.

INTRODUCTION

IN a number of devices an electron or ion is made to cross a gap and to deliver energy to or take energy from an electromagnetic field by a loss or increase of its kinetic energy. Such is the case for the klystron,[1] the reflex klystron,[2] the linear accelerator,[3] etc. In none of these devices may the electron cross the same gap a number of times without getting out of phase with the alternating field and thus the efficiency of such devices is low. In the cyclotron,[4] however, by virtue of the constant angular velocity of charged particles in a magnetic field, the particles do cross the gap many times in phase with the alternating voltage.

THEORY

Let us consider the conditions necessary for an electron to cross a gap many times in phase with the field even though the amplitude of the swing varies over wide limits. This condition is obviously satisfied by simple harmonic motion. For a particle to have this kind of motion the restoring force on it must vary linearly with the distance from the equilibrium or rest position. Thus we write.

$$e(dV/dx) = -Cx \tag{1}$$

V is electrostatic potential, e is the electronic charge, and x is the distance measured from the plane of equilibrium where we place the gap for energy transfer. By integration

$$V = -K^2x^2. \tag{2}$$

A plot of this potential distribution is a parabola.

The period of oscillation may be shown to be

$$P = \frac{2\pi}{K}\left(\frac{m}{2e}\right)^{\frac{1}{2}} \tag{3}$$

which is independent of the amplitude as expected. From this we get an expression for the frequency of oscillation in terms of a, the amplitude, and V, the voltage at $x=a$.

$$f = \frac{1}{2\pi a}\left(\frac{2Ve}{m}\right)^{\frac{1}{2}} \tag{4}$$

or in practical units

$$f = \frac{9.44\times10^6(V)^{\frac{1}{2}}}{a} \tag{5}$$

where a is in cm and V in practical volts. Or we may write

$$V = 1.12\times10^{-14}a^2f^2. \tag{6}$$

If ions are used V is divided by M (the mass ratio of the ion).

It is proposed that devices employing this type of electrical resonance be called Palletrons from the Greek word "pallein" meaning to vibrate and the usual "tron" suffix.

* Presented at the meeting of the American Physical Society in Washington, D. C., May 3, 1947.

[1] R. H. Varian and S. H. Varian, J. App. Phys. **10**, 321 (1939).

[2] W. C. Hahn and G. F. Metcalf, Proc. I.R.E., **27**, 106 (1939).

[3] Wideroe, Archiv f. Elektrotechnil **21**, 387 (1929).

[4] E. O. Lawrence and M. S. Livingston, Phys. Rev. **40**, 19 (1932).

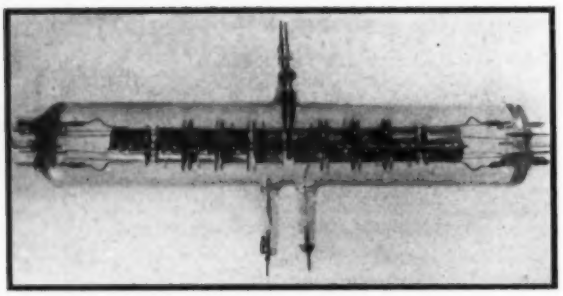


FIG. 1. Photograph of experimental palletron voltage generator. Diameter of cylinders is 2 cm.

APPLICATIONS

There are a number of devices which may be made using this type of electrostatic field, two of which have been constructed and studied experimentally. They are (1) a high voltage generator, and (2) a high frequency oscillator. They confirm the above theory in a precise manner. Their efficiencies are high.

Figure 1 shows the arrangement of elements in an experimental palletron voltage generator. There is an electron emitter in the form of a tungsten filament at the center and the cylindrical elements adjacent to it form a gap across which the radiofrequency voltage is applied. Coupling flanges at the ends of the cylindrical elements augmented by external condensers maintain high capacitive coupling between them so that each half of the device may oscillate in potential as a whole with the applied radiofrequency voltage. Thus the r.f. field appears only at the central gap; inside the cylinders there is only the d.c. parabolic electrostatic field that is provided by a potentiometer or voltage divider for distributing the proper potentials to the cylinder elements. See the circuit schematic of Fig. 2. A paraxial magnetic field, maintained throughout the cylinders by means of Helmholtz coils, constrains the motions of the electrons to the proximity of the axis and prevents them from striking the cylindrical elements. The electron collectors at the ends are shaped like cups so that secondary electrons are trapped and thus eliminated.

The values of resistances shown on Fig. 2 follow a parabolic distribution only approximately. This deviation from the ideal was an experimental expedient used to obtain better operation. Due to the small number of cylindrical elements and to the fact that the line-up in the tube was not perfect, so that some current went to the cylindrical elements, it would not be expected that the potentials applied to the elements for a parabolic potential distribution down the axis would themselves be strictly parabolic. R_5 and R_6 connecting the central rings to ground are unequal because the filament was not centered well in the gap.

Assume that the device is in operation and that the potential of the collectors is that given by Eq. (6). Now consider an electron emitted from the filament when the r.f. field is a maximum. The electron will be accelerated on the positive side and will pass into the cylinder coming to rest at some distance inside and reversing its motion to cross the gap when the r.f. potential is a maximum in the opposite direction. It will be accelerated by the full voltage across the gap and will penetrate to a greater distance in the other cylinder coming to rest, reversing, and arriving at the gap again when the r.f. field is a maximum to receive a further acceleration. This action will be repeated until the electron has acquired enough energy by repeated accelerations to enable it to reach either of the two collector electrodes. It will arrive at this electrode when the r.f. field is substantially zero and with relatively a very small velocity so that the heating of the collector electrode is negligible. This is one reason for the high efficiency. Electrons which start at other times experience the same action but their accelerations on crossing the gap are less so that they must oscillate back and forth a greater number of times before reaching a collector. The steady arrival of electrons at the collector electrodes, cycle after cycle, replenishes the negative charge that is lost by the current drain through the voltage divider resistors and thus maintains the high negative potential.

It remains to be shown that the device is self-starting since initially all of the elements of the cylinders are at zero d.c. potential and there is no parabolic field within them. The lack of this field at the start eliminates the retarding force and the first electrons go clear through the cylinders and reach the collectors driving their potential negative. At all values of collector voltage below that given by Eq. (6) an electron can reach the collector with fewer oscillations so that although

it does not remain in exact phase, nevertheless it acquires sufficient energy to contribute its negative charge to the collector to build up the voltage. Experimentally it has been shown that this is true and the device immediately builds up to the final voltage.

Figure 3 shows the current to the collectors during the starting process. It was obtained by applying negative voltages to the different elements ranging from zero to their final values and measuring the collector current. Actually a resistance network similar to that of Fig. 2 but of lower and strictly parabolic values was used to distribute the proper potentials to the elements and the end or collector voltage was varied from zero upwards. The two collectors were connected together through choke coils so that they could vary with the r.f. potential which was 140 volts peak across the gap. Note that the current is never zero at any point in the build-up process. The maximum of the curve at 480 volts corresponds to the value obtained from Eq. (6) for the frequency used (24 megacycles). This is also the voltage that was built up with the device operating in the usual way at this frequency.

With a peak voltage of 140 applied across the gap at a frequency of 38 megacycles, the measured collector voltage was 1338 volts. This means that the minimum number of transits across the gap for an electron was nine and one half. The magnetic field strength was 157 gauss.

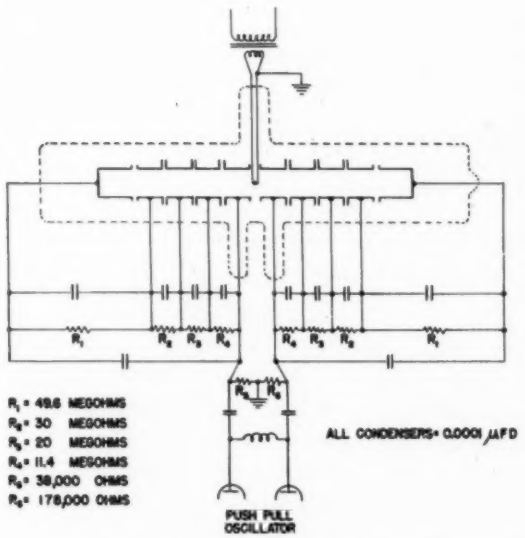

R1 = 49.8 MEGOHMS
R2 = 30 MEGOHMS
R3 = 20 MEGOHMS
R4 = 11.4 MEGOHMS
R5 = 38,000 OHMS
R6 = 178,000 OHMS
ALL CONDENSERS = 0.0001 μF D
PUSH PULL OSCILLATOR

FIG. 2. Diagram of circuit used with experimental tube.

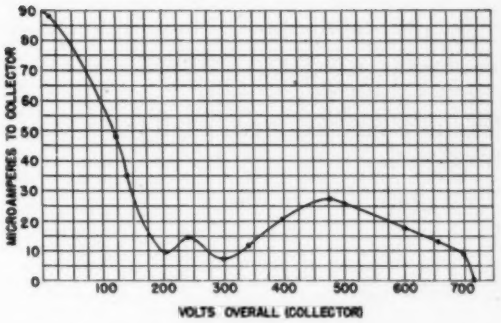


FIG. 3. Collector current during the starting-up period.

It appears that this device might extend the range of voltages that can be generated since all of the high voltage elements are enclosed in the vacuum and corona losses are thus eliminated. Field emission toward the ends of the cylinder where the potential gradient is high is probably the deciding factor in determining the value of "a" and hence the overall dimension.

A palletron generator for a million volts with a half working length "a" equal to one meter would operate at a frequency of 94.4 megacycles if we assume a purely parabolic distribution. However, because of the relativistic increase in mass of the high velocity electrons they would slow up and get out of phase with those just starting out. The solution lies in deviating from the parabolic potential distribution, this deviation being such as to decrease the period of the electrons. This deviation would be negligible near the center and would increase toward the outer ends of the device. It is proposed that the minimum number of gap transits for the electrons would be 100 and thus the applied radio frequency would need to reach 10,000 volts peak across the gap. It is further proposed to produce the magnetic field by means of Alnico ring magnets incorporated in the cylindrical elements of which there might be 100 in. each half. This number would reduce the voltage between any two low enough to prevent field emission. The resistances for the voltage divider would also be incorporated in the tube as spacers between the cylindrical elements.

ACKNOWLEDGMENTS

The first palletron voltage generator was built and operated successfully at the Bell Telephone

Laboratories just prior to the war. Publication was prevented by the fact that the patent application[5] was put in the secret category. This restriction is now removed and the patent has been issued. The tube herein described was made and tested at the National Union Research Laboratories.

The writer wishes to acknowledge the able assistance of Mr. H. J. Koch and Mr. W. Hostetler in this work.

[5] A. M. Skellett, U. S. Patent No. 2,407,298, Filed Dec. 15, 1942; Issued Sept. 10, 1946.

Measurement of Stress by Means of X-Rays

D. E. Thomas
Armament Research Department, Woolwich, England
(Received September 11, 1947)

The article shows how the formulae used in stress measurement by x-rays are derived and points out an error in a recently published book.

A RECENT book by Sproull[1] contains references to techniques developed by the present author for the measurement of stress or strain in metals by means of x-rays. Equations (21–26) and (21–28) in the book are attributed to the author of this article. The former equation is quoted correctly but the latter is not. The difference between Sproull's formula (21–28) and the author's as published[2] is only that he gives "$\sec^2 2\eta$" as a factor instead of "$\sec 2\eta$." In all cases in the author's experience of the use of this formula, η, the complement of the Bragg angle, is of the order of 9.5° (not 5° as stated by Sproull) so that 2η is about 19° and $\sec 2\eta$ is 1.058. Thus, the difference between a stress calculated by Sproull's formula and by the present author's is less than 6 percent. If the value of η were 5°, $\sec 2\eta$ would be only 1.015 and the discrepancy would be much less than the experimental error of measurement; but 6 percent is too near the order of the experimental error to be neglected.

In the paper quoted,[2] owing to editorial restriction, it was not possible to show how the formulae quoted were derived. The present appears to be an opportune moment to do so.

Stress, or rather, strain measurement by x-rays is really a form of precision measurement of the crystal lattice. From the Bragg equation,

$$\lambda = 2d\sin\theta;$$

$$\frac{\partial\theta}{\partial d} = -\frac{1}{d}\tan\theta,$$

so that the change of θ with change of lattice dimension is proportional to $\tan\theta$ and so increases very rapidly as θ approaches 90°. This, of course, means that back reflections are the most sensitive experimental arrangements for this technique.

If S_1 and S_2 are the principal stresses in a body (Fig. 1), the extension $\epsilon_{\phi\psi}$ in the direction $\phi\psi(OP)$ is given by:

$$E\epsilon_{\phi\psi} = S_1[\cos^2\phi\sin^2\psi - \nu\sin^2\phi\sin^2\psi - \nu\cos^2\psi] + S_2[\sin^2\phi\sin^2\psi - \nu\cos^2\phi\sin^2\psi - \nu\cos^2\psi],$$

which may be reduced to

$$E\epsilon_{\phi\psi} = (1+\nu)\sin^2\psi S_\phi - \nu(S_1+S_2),$$

where S_ϕ is the stress along ON and

$$S_\phi = S_1\cos^2\phi + S_2\sin^2\phi;$$

E is Young's modulus and ν is Poisson's ratio.

If x-rays of wave-length λ are incident on a strained surface at O (Fig. 2), the reflections at F_1 and F_2 on a film normal to XO will be produced by planes whose normals are OZ_1 and OZ_2, and these normals are inclined to the normal to the surface of the body at angles $(\alpha+\eta)$ and $(\alpha-\eta)$. The extensions corresponding to

[1] Wayne T. Sproull, *X-Rays in Practice* (McGraw-Hill Book Company, Inc., New York, 1946).
[2] D. E. Thomas, J. Sci. Inst. **18**, 135 (1941).

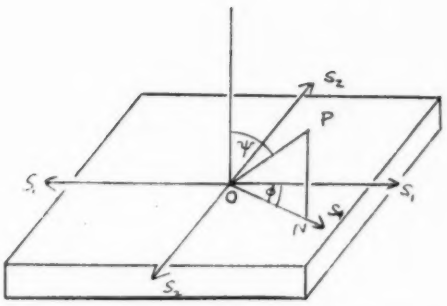


FIG. 1. Illustrating directions of stresses.

these reflections at F_1 and F_2 are:

$$\epsilon_1=\frac{S_\phi}{E}[(1+\nu)\sin^2(\alpha+\eta)]-\frac{\nu(S_1+S_2)}{E},$$

$$\epsilon_2=\frac{S_\phi}{E}[(1+\nu)\sin^2(\alpha-\eta)]-\frac{\nu(S_1+S_2)}{E}.$$

The distance apart of the two reflections F_1 and F_2 is

$$D(\tan 2\eta_1+\tan 2\eta_2)=W.$$

Hence, the change in this distance due to strain is

$$\partial W=D[(\partial/\partial d)\tan 2\eta\partial d_1+(\partial/\partial d)\tan 2\eta\partial d_2].$$

Now, ∂d_1, ∂d_2 are the changes in d on the two sides of the incident ray, and $\partial d_1=\epsilon_1 d_0$ and $\partial d_2=\epsilon_2 d_0$, where d_0 is the unstrained value of the spacing. Then

$$\partial W=Dd_0(\epsilon_1+\epsilon_2)\sec^2 2\eta 2(\partial\eta/\partial d).$$

Now

$$\lambda=2d\sin\theta=2d\cos\eta;$$

hence,

$$\cos\eta-d\sin\eta(\partial\eta/\partial d)=0$$

or

$$\partial\eta/\partial d=(1/d)\cot\eta,$$

so that

$$\partial W=2D(\epsilon_1+\epsilon_2)\sec^2 2\eta\cot\eta.$$

Now

$$\epsilon_1+\epsilon_2=(S_\phi/E)(1+\nu)\{\sin^2(\alpha+\eta)+\sin^2(\alpha-\eta)\}-(2\nu/E)(S_1+S_2),$$

and the expression $\{\sin^2(\alpha+\eta)+\sin^2(\alpha-\eta)\}$ can be shown to be equal to

$$2(\sin^2\alpha\cos 2\eta+\sin^2\eta).$$

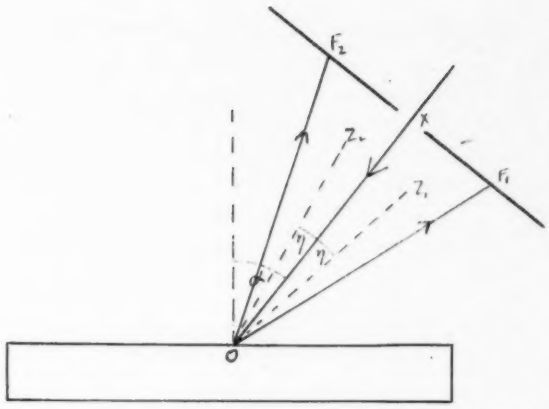


FIG. 2. Reflections from stressed surface with oblique incidence.

Hence,

$$\partial W=4D\sec^2 2\eta\cot\eta\left[\frac{S_\phi}{E}(1+\nu)\times(\sin^2\alpha\cos 2\eta+\sin^2\eta)-\frac{\nu}{E}(S_1+S_2)\right].$$

This is the change in the diameter of the back reflection due to strain. We cannot measure this since we do not know the unstrained diameter. But if we measure the diameter for another value of the angle of incidence, say α', the change in diameter due to strain in this case will be

$$\partial W'=4D\sec^2 2\eta\cot\eta\left[\frac{S_\phi}{E}(1+\nu)\times(\sin^2\alpha'\cos 2\eta+\sin^2\eta)-\frac{\nu}{E}(S_1+S_2)\right].$$

The difference in diameter measured at the two angles of incidence α and α' is then

$$\partial W-\partial W'=4D(S_\phi/E)(1+\nu)\sec^2 2\eta\times\cot\eta\cos 2\eta(\sin^2\alpha-\sin^2\alpha')=4D(S_\phi/E)(1+\nu)\sec 2\eta\times\cot\eta(\sin^2\alpha-\sin^2\alpha').$$

This is the equation of what Sproull calls "Thomas's technique B." Sproull's calculation really involves putting the expression

$$\{\sin^2(\alpha+\eta)+\sin^2(\alpha-\eta)\}\text{ equal to }2\sin^2\alpha.$$

Then

$$\partial W=4D\sec^2 2\eta\cot\eta[(S_\phi/E)(1+\nu)\sin^2\alpha-(\nu/E)(S_1+S_2)],$$

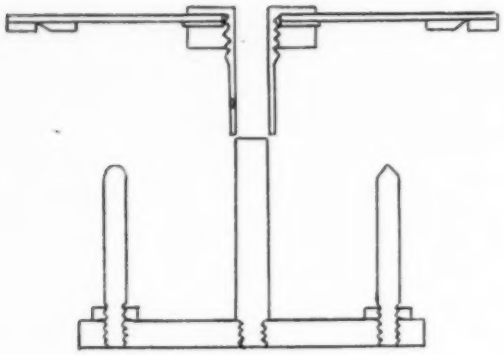


FIG. 3. The film pricker.

and

$$\partial W' = 4D\sec^2 2\eta \cot\eta [(S_\phi/E)(1+\nu)\sin^2\alpha' - (\nu/E)(S_1+S_2)],$$

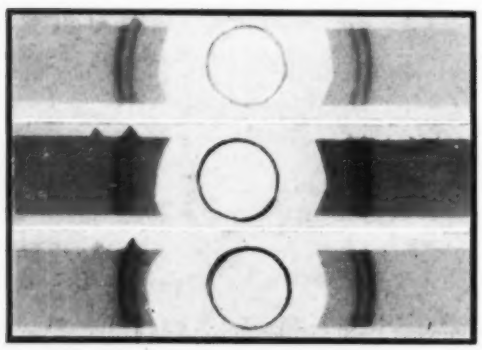
so that

$$\partial W - \partial W' = 4D(S_\phi/E)(1+\nu)\sec^2 2\eta \cot\eta(\sin^2\alpha - \sin^2\alpha').$$

This is Sproull's Eq. (21–28). The assumption made is that the value of the elastic extension in the direction of the incident ray is the mean of the values in the directions OZ_1 and OZ_2.

What Sproull calls "Thomas's technique A" is a technique for measuring the stress in one back reflection photograph, in which the difference in the distance of F_1 and F_2 from X is measured (Fig. 2). The relevant equation is deduced as follows: In Fig. 2,

$$XF_1 = D\tan 2\eta_1 \quad \text{and} \quad XF_2 = D\tan 2\eta_2,$$



(a)

(b)

(c)

FIG. 4. Back reflections from mild steel: (a) 5 minutes exposure, 2-mm diameter front aperture. (b) 1 minute exposure, 2-mm diameter front aperture. (c) 0.5 minute exposure, 4-mm diameter front aperture.

so that

$$\begin{aligned} XF_1 - XF_2 &= D(\tan 2\eta_1 - \tan 2\eta_2) \\ &= D(\partial/\partial d)\{\tan 2\eta d_0(\epsilon_1 - \epsilon_2)\} \\ &= D\sec^2 2\eta 2(\partial\eta/\partial d) d_0(\epsilon_1 - \epsilon_2) \\ &= 2D\sec^2 2\eta \cot\eta(\epsilon_1 - \epsilon_2). \end{aligned}$$

Now

$$\begin{aligned} \epsilon_1 - \epsilon_2 &= (S_\phi/E)(1+\nu)\{\sin^2(\alpha+\eta) - \sin^2(\alpha-\eta)\} \\ &= (S_\phi/E)(1+\nu)4\sin\alpha\cos\alpha\sin\eta\cos\eta. \end{aligned}$$

So that

$$\begin{aligned} XF_1 - XF_2 &= 4D(S_\phi/E)(1+\nu)\sin 2\alpha \times \sec^2 2\eta \cot\eta \sin\eta\cos\eta \\ &= 4D(S_\phi/E)(1\mp\nu)\sec^2 2\eta \times \cos^2\eta \sin 2\alpha. \end{aligned}$$

In order to use this technique one must know where X is on the film. But a hole is punched through the film around this position to pass the incident radiation. This difficulty is, however, easily surmounted by the use of the tool shown in Fig. 3. (This tool is an improvement on the one described in the paper mentioned above.) The film holder is fitted with a central bushing which is an accurate fit on the center leg of the tool shown. It is also an accurate fit on the collimating tube of the x-ray tube. One of the outer legs of the tool of Fig. 3 is turned to a 90° point, and the film holder is allowed to slide down the central leg until the film rests on the two outer legs. Gentle pressure on the point will then make a fine mark on the film. The film is now turned through 180° and another mark made. These two marks are equidistant from the axis of the bushing of the film holder. If the incident ray had been positioned on the axis of this bushing, we should have all the information we required to measure $XF_1 - XF_2$ and so to calculate the stress. But experience has taught that the incident ray is sometimes eccentric to the collimating tube. This condition is readily detected and corrected. If a reflection is taken from a symmetrical lattice, an unstrained body, or even a normal reflection from a strained body, and the two reflections are not found to be equidistant from the pricker marks, the incident beam emerging from the x-ray tube is not coaxial with the collimating tube. The difference between these two distances can be measured and applied as a correction to all subsequent measurements.

Technique B, using a distance piece to give the film-specimen distance—(see Fig. 2)—is capable of measuring a really good lattice, e.g. iron filings heat treated in vacuum, with a probable error of only one part in 100,000, all precautions for accuracy being observed. This is the vindication of the technique. Lattices of industrial steels are always found to vary by much more than this, even to one part in 10,000. This must be the fault of the steel and not of the technique. These variations of lattice dimensions are actually present in the steel and are indications of internal stress.

Technique B is generally practiced in two exposures, one at normal incidence and the other with incidence at 30° or 45° to the normal. Technique A is practiced in one exposure with incidence at 30° to 45° to the normal. Technique *A* does not provide such an accurate measurement as technique B as the difference in the two stresses being compared is much less in the former. But it does measure the stress in one direction in one exposure and the two reflections being used are certainly produced at the same point of the specimen. There is always a small element of doubt on this point in technique B.

Dr. Sproull doubts the author's statement that satisfactory photographs can be obtained in five minutes with a front pinhole of 2-mm diameter, using 20 ma and 30 kv. In Fig. 4, photographs are shown which have been obtained under these and harder conditions from a piece of mild steel such as is used in shipbuilding. Figure 4(a) was given 5 minutes exposure, Fig. 4(b) one minute, both using a 2-mm diameter front aperture and Fig. 4(c) was given half a minute with a front aperture of 4-mm diameter. The photographs were all taken at normal incidence with a film-specimen distance of 50 mm, using 20 ma and 30 kv. The focusing conditions were, of course, carefully considered and satisfied, and the film was rotated throughout the exposure.

Permission to publish this paper has been granted by the Chief Scientist, Minister of Supply.

Measured Impedances of Helical Beam Antennas

Otto J. Glasser and John D. Kraus
Communications Laboratory, Department of Electrical Engineering, Ohio State University, Columbus, Ohio
(Received September 26, 1947)

Experimentally determined impedance characteristics are presented for helical antennas. At frequencies too low for the axial or beam mode, the terminal impedance changes greatly with small variations in frequency. However, in the frequency range of the fundamental axial mode, the impedance is relatively constant, especially if the number of turns and the pitch angle of the helix are not too small. This characteristic is an advantage in wide frequency-band applications. For example, the resistance variation of a 5-turn, 18-degree helix over a frequency band of about 1.7 to 1 is only 105 to 145 ohms, and the reactance variation only 0±15 ohms. The effect of the pitch angle and number of turns is discussed.

INTRODUCTION

It has been shown that a helix with a circumference of the order of one wave-length can be excited as a beam antenna.[1] Radiation is maximum in the direction of the helix axis and is circularly polarized or nearly so. This mode of radiation is called the axial or beam mode. The important characteristics of a helical antenna, with the exception of its impedance, are reported in another paper.[2] In most antenna applications a knowledge of the terminal impedance is necessary, and it is the purpose of this article to present impedance data for uniform helical antennas of circular cross section.

Referring to Fig. 1, the following symbols are used: D = diameter of helix cylinder, S = spacing between turns, and α = pitch angle = $\arctan(S/\pi D)$.

[1] J. D. Kraus, "Helical beam antenna," Electronics **20**, No. 4 109 (1947).

[2] J. D. Kraus and J. C. Williamson, "Characteristics of helical antennas radiating in the axial mode," J. App. Phys. **19**, 87 (1948)

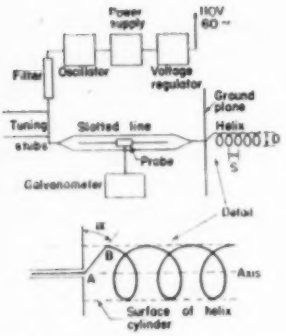


FIG. 1. Equipment for impedance measurements.

The general arrangement of the equipment for the impedance measurements is illustrated in Fig. 1. The helix is supported on a 1.5×1.5 meter copper ground plane, as shown by Fig. 2. Mechanical support is provided by 3 plastic tubes and a thin plywood ring at the end. The helix is energized by a 53-ohm coaxial transmission line, with the inner conductor connected to the helix and the outer conductor terminated in the ground plane, as illustrated by the detail in Fig. 1. Beyond point B, the helical conductor lies in the surface of the imaginary cylinder of the helix. Between terminal A, at the end of the coaxial line, and point B, the conductor lies in a plane through the helix axis and at approximately the same pitch angle, α, as for the helix proper.

FIG. 2. 6-degree helix mounted on ground plane for impedance measurements.

MEASUREMENTS

The terminal impedance of helical antennas was measured as a function of 3 variables: frequency, pitch angle (α), and number of turns (n). The impedance of the helix was measured in each case as a function of frequency, and two sets of measurements were made: (1) with the physical length constant but with different values of pitch angle and (2) with the pitch angle constant and different numbers of turns. The results are shown by Figs. 3 and 5. In Fig. 3 the impedance is presented as a function of frequency for 4 helices of about the same physical diameter ($D = 22.5$ cm) and axial length from ground plane (123 cm) but of different pitch angles. The frequency in Mc/sec. is shown at points along the impedance spirals. The 4 helices have pitch angles of 6, 12, 18, and 24 degrees and 15, 8, 5, and 3.9 turns, respectively. All helices are of uniform circular cross section and are wound of 1.27-cm diameter copper tubing. These helices are the same as used for the pattern measurements of reference 2 (Figs. 9 and 10). In the first set of measurements both α and n are changed, but as shown by the second set, the effect of n is small for 12-degree helices, if the number of turns is sufficient, and is probably small for other pitch angles. In applications there are usually definite limitations as to the physical size of an antenna, so that maintaining the axial length constant while varying α (and also n) has practical significance.

Effect of Frequency

Referring to Fig. 3, it is interesting to note that the impedance-frequency behavior is distinctly different at fundamental axial mode frequencies and at frequencies which are too low. Based on pattern measurements (reference 2), the fundamental axial mode* for the 12- and 18-degree helices is between about 300 and 500 Mc. For the 6-degree helix the fundamental axial

* The axial or beam mode of radiation of a helical antenna may be arbitrarily related to the form of the radiation pattern. As mentioned in reference 2, the axial mode is characterized by a single major lobe in the direction of the helix axis with nearly circular polarization. Actually this mode can be described as the fundamental or lowest order axial mode. At frequencies too high for the fundamental mode, the pattern, in general, disintegrates into conical or multilobed forms, and these patterns can be considered as belonging to higher order axial modes.

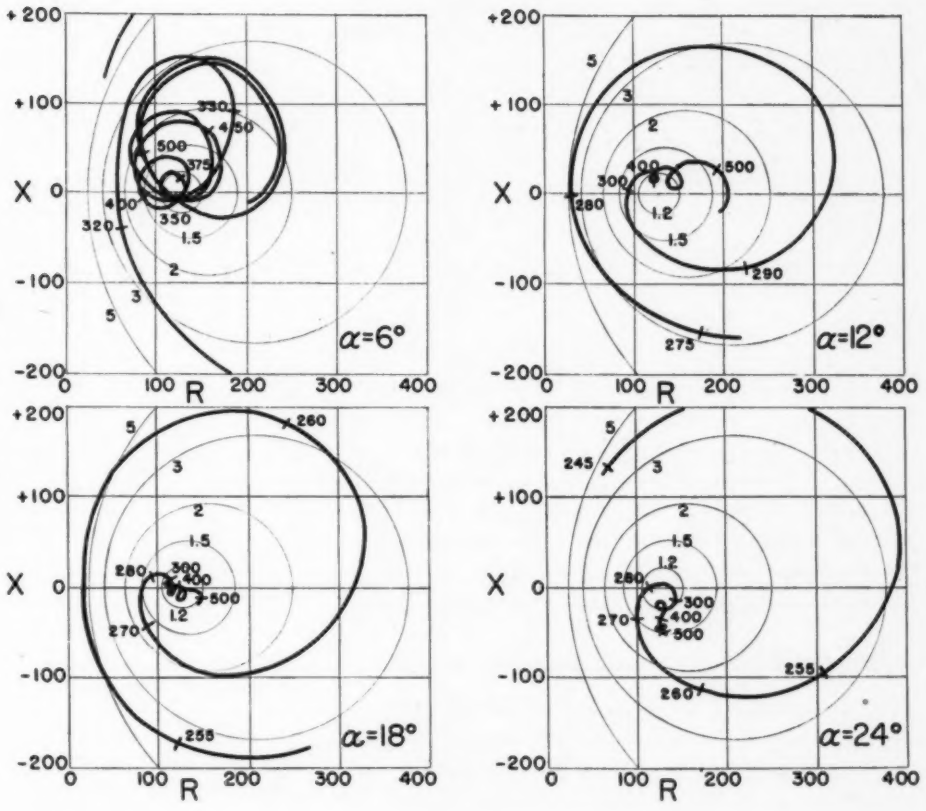


FIG. 3. Ohms resistance, R, and ohms reactance, X, of the terminal impedance as a function of frequency in megacycles for four helices of the same physical diameter and length but having pitch angles of 6, 12, 18, and 24 degrees. The standing wave ratio on a 125-ohm line is indicated by comparison with the circles.

mode occurs between about 350 and 450 Mc, while no well defined fundamental axial mode pattern was found for the 24-degree helix, although one may be present over a narrow frequency band. Referring in particular to the 12- and 18-degree helices, it is apparent that at frequencies too low for the axial mode (<300 Mc) the impedance changes greatly with small variations in frequency. This variation is typical of a long transmission line with a nearly pure standing wave. The impedance goes through a complete cycle of values (once around the impedance diagram) for a frequency difference given approximately by

$$\Delta f = c/2nL,$$

where Δf = frequency difference in c.p.s., $c = 3 \times 10^8$ meter/sec., L = length of one turn of helical conductor in meters, and n = number of turns.

On the other hand, at frequencies in the range of the fundamental axial mode, the impedance variation of the helices is relatively small, except for the 6-degree helix. The constancy of the impedance indicates that the wave which is reflected from the open end and travels back to the terminals is small. This behavior is similar to that of a transmission line terminated in approximately its characteristic impedance.

The standing wave ratio of voltage (SWR) which would be present on a 125-ohm line connected to the helix is indicated by comparison with the circles shown in Fig. 3. It appears that the SWR for the 18-degree helix is less than 1.25 over a frequency range of more than 300 to 500 Mc. This is illustrated more clearly by the SWR *vs.* frequency diagram in Fig. 4. The SWR is less than 1.25 over a frequency band width of about 80 percent and less than 1.1 over a band width of about 20 percent. The terminal imped-

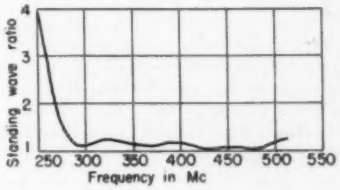


FIG. 4. Standing wave ratio as a function of frequency on a 125-ohm transmission line for 5-turn, 18-degree helix.

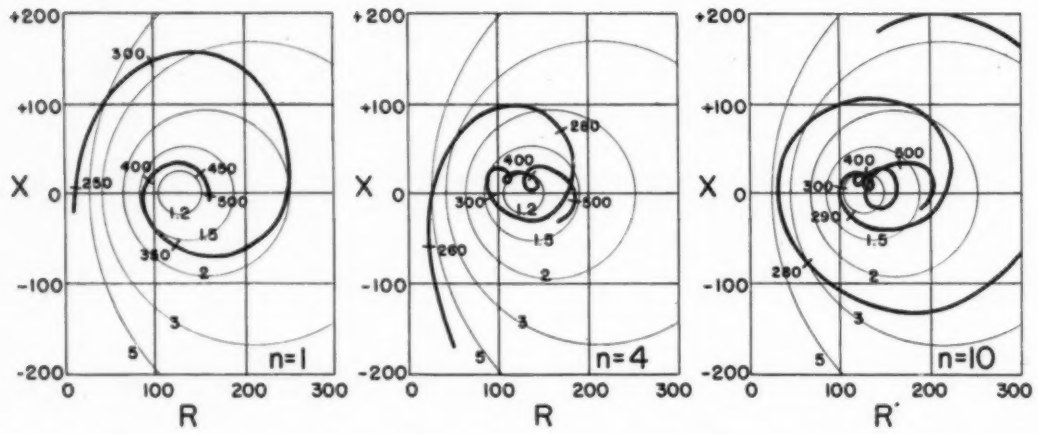


FIG. 5. Ohms resistance, R, and ohms reactance, X, of the terminal impedance as a function of frequency in megacycles for 12-degree helices of 1, 4, and 10 turns showing effect of number of turns.

ance variation of the 18-degree, 5-turn helix in the frequency range 285 to 500 Mc is 105- to 145-ohms resistance and 0±15-ohms reactance.

Effect of Pitch Angle and Number of Turns

It is apparent from Fig. 3 that the impedance variation between 300 and 500 Mc decreases with increase in pitch angle. However, the geometric mean resistance** in this frequency range is about 130 ohms for all except the 6-degree helix.

The effect of the number of turns on the terminal impedance is shown by Fig. 5. The measured impedance variation as a function of frequency is shown for a 12-degree helix (D = 23.2 cm) of 1, 4, and 10 turns wound of 1.27-cm diameter copper tubing. Patterns for helices of these dimensions are shown by Fig. 11 of reference 2. The difference in the impedance behavior of the 4- and 10-turn helices in the vicinity of 400 Mc is small and suggests that if the number of turns is sufficient (say, $n>3$) the impedance is relatively independent of n for axial mode frequencies. It may also be noted that the geometric mean resistance between 300 and 500 Mc is about 127 ohms for both the 4- and 10-turn helices.

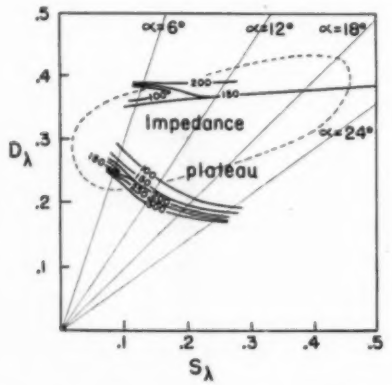


FIG. 6. Map of impedance as a function of the turn spacing (S_λ) and helix diameter (D_λ) in wave-lengths. The solid contours give the magnitude of the terminal impedance in ohms. The dashed contour shows the approximate limits of the fundamental axial mode as determined by pattern measurements.

Effect of Diameter and Spacing

The terminal impedance as a function of the turn spacing and helix diameter in wave-lengths is presented in Fig. 6, the data being taken from Fig. 3. The solid contours give the magnitude of the impedance in ohms. The contour interval is 50 ohms. This map reveals a broad impedance plateau between the 100- and 150-ohm contours and shows that the impedance is relatively constant over a wide range of helix dimensions. Beyond the upper and lower limits of the plateau the impedance changes rapidly with the diameter in wave-lengths (D_λ) as indicated by the close bunching of the contours. The dashed contour is the 50-degree beam-width contour from Fig. 9 of reference 2 and shows the approximate limits of the fundamental axial mode. It is apparent that the fundamental axial mode occurs over a large part of the impedance plateau.

** The geometric mean resistance equals $(R_1R_2)^{\frac{1}{2}}$, where R_1 and R_2 are the maximum and minimum resistance values.

Effect of Other Parameters

All of the measurements reported above are on helices with a conductor diameter of 1.27 cm or about 0.02 wave-length at the frequencies employed. The conductor diameter is a factor in determining the helix impedance. In the measurements described the frequency is varied so that the pitch angle is preserved, but the helix diameter and conductor diameter in wave-lengths change. If the conductor diameter is changed by a large factor it will produce considerable effect. In the measurements reported, the conductor diameter in wave-lengths changed as a function of the frequency by about 2 to 1, and the effect of this change is probably small compared to the effect of the 2 to 1 change in D_λ. Hence, the measurements should be quite characteristic of helices with a conductor diameter of the order of 0.02 wave-length.

The size and shape of the ground plane can produce an effect on the measured impedance. The data presented herein were measured with a 1.5×1.5 meter ground plane or about 2×2 wave-lengths at the frequencies employed. Provided that the ground plane is not too small, the effect of size and shape will be to produce differences in detail although not in the general form of the impedance curves.

The dielectric structure supporting the helix also influences the measured impedance. By way of comparison, some measurements on helices in the almost complete absence of dielectric material, except air, showed that the amount used for mechanical support has some effect on the detail but not on the general form of the impedance curves.

CONCLUSIONS

At frequencies too low for the axial mode the impedance varies rapidly with frequency, but in the frequency range of the axial mode the impedance is relatively constant. For helices of fixed physical length, the impedance variation with frequency decreases with increase in pitch angle. The effect of the number of turns appears to be small at fundamental axial mode frequencies if n exceeds a certain minimum value.

Since the impedance is relatively constant over a wide frequency band and since the pattern and polarization characteristics are also suitable over the same band, the helical beam antenna is naturally fitted for wide frequency-band applications.

Dark-Field Electron Microscopy. I. Studies of Crystalline Substances in Dark-Field

C. E. Hall
Massachusetts Institute of Technology, Cambridge, Massachusetts
(Received September 18, 1947)

Dark-field images in the electron microscope were studied by means of an objective aperture system fixed to the object rather than to the objective lens. The resolution in images produced by the diffusely scattered component is in the range 100 to 200A, but the resolution in images produced by Bragg reflections approaches that obtainable in bright-field operation. A resolution of about 50A was obtained with test objects consisting of evaporated films containing small crystallites, but it is estimated that crystallites having dimensions down to about 20A can be recorded with conditions prevailing in this work. Substances studied in dark-field include evaporated films of metals and compounds and finely divided substances. It is concluded that the dark-field method is of value for studying the size, location, and other characteristics of crystalline components in a specimen. Some of the theoretical factors influencing resolution in dark-field are discussed.

1. INTRODUCTION

IN electron microscopy, as in light microscopy, a dark-field image is one for which the rays which would reach the image plane without deviation by the object are intercepted by an opaque mask. Several schemes for securing this condition in the electron microscope have been described,[1,2] three of which are shown in Fig. 1. The first, 1a, consists of an annular stop in the condenser lens and a circular opening in the objective lens. In the second method, 1b, a conventional circular diaphragm is used in the objective, but the illuminating system is tilted so that the undeviated rays strike the edge of the diaphragm. The third scheme shown in Fig. 1c employs an eccentric objective aperture to intercept the transmitted rays and transmit some of the scattered rays.

It is difficult to say when the possibility of using dark-field illumination in electron microscopy was first recognized. Probably it was obvious to the early workers in the field from the analogy with light microscopy. Certainly, every electron microscopist who has operated an electron microscope with a small objective aperture has seen a dark-field image of some sort, since constant attention is necessary to keep the instrument from lapsing through misalignment into conditions b or c of Fig. 1. Even when the instrument is properly aligned for bright-field operation, the peripheral portions of the intermediate image are seen in dark-field when a small objective aperture is present. The first published dark-field micrographs appear to be those of Boersch,[3,4] who in 1936 showed several dark-field micrographs of gold foils. In these experiments an eccentric aperture was used corresponding roughly to Fig. 1c and although the magnifications are only 60×, the essential principles of dark-field electron microscopy are contained in these two papers. Later, Boersch[5] showed low magnification, dark-field images of gas jets such as are used in electron diffraction experiments with gases. It is therefore apparent that when the electron microscope was in an early stage of development the possibilities and feasibility of dark-field electron microscopy were known.

There appears to have been little interest in dark-field electron microscopy until 1940, when von Ardenne[6] described his "universal electron microscope" in which provision was made for introducing suitable apertures into the objective lens to obtain dark-field micrographs. Several micrographs have been shown made with this instrument to demonstrate the practicality of the method.[2] At about the same time Borries and Ruska[7] published a series of dark-field and bright-field micrographs made at three different ac-

[1] V. K. Zworykin, G. A. Morton, E. G. Ramberg, J. Hiller, and A. W. Vance, *Electron Optics and the Electron Microscope* (John Wiley and Sons, Inc., New York, 1945).
[2] M. von Ardenne, *Elektronen Übermikroskopie* (Verlag Julius Springer, Berlin, 1940).
[3] H. Boersch, Ann. d. Physik **26**, 631 (1936).
[4] H. Boersch, Ann. d. Physik **27**, 75 (1936).
[5] H. Boersch, Zeits. f. Physik **107**, 493 (1937).
[6] M. von Ardenne, Zeits. f. Physik **115**, 339 (1940).
[7] B. von Borries and E. Ruska, Zeits. f. Physik **116**, 249 (1940).

celerating voltages. The micrographs were from the same part of the specimen and show qualitatively the dependence of scattering and image contrast on accelerating voltage. In a study of contour phenomena, Boersch[8] has made use of the dark-field method to study the scattered intensity from the edges of relatively opaque objects.

In the American literature, only three publications have been located in which dark-field micrographs were reproduced although possibly others exist. One occurs in a review article by Anderson.[9] Another is a note by Levy,[10] in which he demonstrates the feasibility of securing dark-field micrographs with the RCA type B instrument by tilting the illuminating system in the manner shown in Fig. 1b. In a note describing the multiple aperture system used in this work, the author included a dark-field micrograph showing features not visible in bright-field and attributed to Bragg reflections. On the theoretical side, Schiff[12,13] has reported the results of calculations regarding the possibility of recording individual atoms by the dark-field method.

Although the possibility of using dark-field illumination in electron microscopy has long been recognized, the method has attracted little attention and has been neglected in comparison to the other aspects of the subject. One factor which tends to discourage investigations in dark-field is the apparently troublesome technique involved. By means of the multiple aperture system used in this work, dark-field micrographs may be made quickly and with little effort and without disturbing the adjustments necessary for bright-field operation.

2. APPARATUS AND METHODS

All micrographs were made with an RCA type B electron microscope with the accelerating voltage raised to 65,000 volts by the insertion of an additional 45-volt battery in the high voltage stabilizer circuit. Except for the aperture system, electron gun, and pole pieces, the instrument is conventional. The pole pieces (constructed in this laboratory) have a focal length of 3.5 mm and were "symmetrized" after the method of Hillier and Ramberg.[14] For most of the work, the bright-field resolution of the instrument was in the range 25 to 30A for test objects containing heavy metal particles such as gold. Micrographs were recorded at from 10,000× to 20,000×, and bright-field micrographs were usually made without an objective aperture.

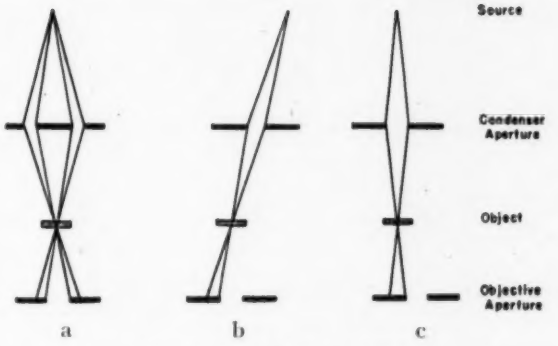


FIG. 1. Dark-field methods.

In dark-field operation, an objective aperture system was used which requires no alteration to the electron microscope since it is suspended from the specimen holder.[11] The device consists essentially of a second screen mounted below the specimen screen, so that for enhanced contrast in bright-field operation, each opening of the specimen screen has below it a corresponding opening serving as an aperture stop. In dark-field operation, the meshes are slightly misaligned so that a part of each object is masked by a solid portion of the aperture mesh. This condition is equivalent to Fig. 1c. The chief advantage of the device lies in the rapidity with which apertures can be removed and the instrument converted from dark-field to bright-field operation. There are, however, some disadvantages. It is difficult to obtain micrographs of the same place in both dark- and bright-field, since the object must be removed from the instrument to make the conversion. Whether a particular portion of the specimen appears in dark-field is largely a matter of accident. Although this is a serious disadvantage if a particular structure is being sought, it is of little consequence if the specimen is uni-

[8] H. Boersch, Physik. Zeits. **44**, 32 (1943).
[9] T. F. Anderson, in E. O. Kraemer, ed., *Advances in Colloid Science* (Interscience, New York, 1942), Vol. 1, pp. 353-390.
[10] G. B. Levy, J. App. Phys. **15**, 623 (1944).
[11] C. E. Hall, J. App. Phys. **18**, 588 (1947).
[12] L. I. Schiff, Phys. Rev. **61**, 391 (1942).
[13] L. I. Schiff, Phys. Rev. **61**, 721 (1942).
[14] J. Hillier and E. G. Ramberg, J. App. Phys. **18**, 48 (1947).

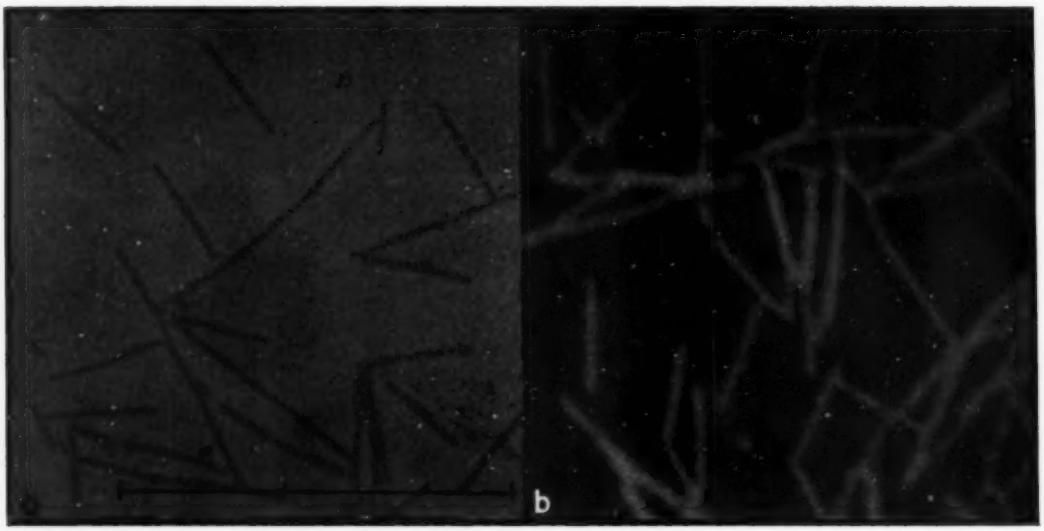


FIG. 2. Tobacco mosaic virus, a, bright-field, b, dark-field.

form, as was the case in most of this work. With a distance of 3 mm between meshes as originally used, a weak illumination sometimes reached the image plane through openings other than the one directly under the point of observation. Later, the distance was reduced to 1.5 mm to reduce stray illumination and simplify the alignment procedure. A single, circular aperture was sometimes used in place of the aperture mesh, but in this case only about four object openings could be observed without removing the object from the microscope and moving the aperture.

Since the scattered intensity is much less than the incident, a higher illumination is required for satisfactory dark-field operation. Since dark-field images are faint with the conventional unbiased gun, the self-biasing arrangement described by Hillier[15] was installed, which provides adequate intensity for observation in dark-field at magnifications up to 20,000×. Exposure times are about 5 seconds.

In all instances where the thickness of evaporated films is given below, values have been obtained on the assumption that material evaporated from a tungsten filament is radiated with spherical symmetry. The assumption is, of course, open to question. It has been found, at least, that the results are reasonably reproducible and presumably could be repeated by other workers. In all cases, the specimens were supported directly above the tungsten filament which was wound in a helical spiral. Filament-to-object distances varied from 6 to 13 cm.

The conditions of dark-field observation present special problems in connection with the supporting film. It is desirable that the film have negligible scattering power per unit area compared to the objects to be examined and that it be capable of withstanding the heat produced by the high intensity electron beam. Collodion and other plastics are not entirely suitable because they cannot be made thinner than about 75A (estimated from the weight of films cast on water) and because they are weakened by the heat generated by the beam. Films of silicon, beryllium, silicon dioxide (quartz), and silicon monoxide were tried. Films were made by evaporating in vacuum measured amounts of these materials on 125A collodion films supported on the usual 200-per-inch Lektromesh screens. The collodion was subsequently dissolved in acetone. Silicon is remarkably strong and resistant to chemicals, but is difficult to evaporate in measured amounts.[16] Beryllium has very low scattering power and can be made in calculated thicknesses of the order of 20A and less, but the films crystallize and show structure in dark-field. Quartz films were

[15] J. Hillier, J. App. Phys. **17**, 307 (1946).

[16] L. O. Olsen, C. S. Smith, and E. C. Crittendon, Jr., J. App. Phys. **16**, 425 (1945).

found to be quite satisfactory in calculated thicknesses of about 20A. Although the scattering power is a little greater than that of beryllium, quartz shows negligible structure in dark-field. Toward the latter part of the work, a sample of silicon monoxide* was obtained which was found to be superior to quartz since it evaporates at a lower temperature. The existence of this substance seems not to be widely known at present and was suggested as a film material by Professor R. C. Williams, University of Michigan. The low melting point is desirable because it makes for ease of evaporation and reduces the chance of evaporating tungsten from the filament along with the film material.

3. OBSERVATIONS IN DARK-FIELD

I. General Features of Dark-Field Images

It is convenient to divide dark-field images into two categories, those produced by diffuse scattering and those produced by crystalline diffractions. In general, diffuse scattering corresponds to the central halo and background of an electron diffraction diagram while diffraction images correspond to the diffraction rings or spots. There is, however, no sharp distinction between the two types of dark-field images. It is to be expected that there will be a gradual transition and it may not always be possible to say, definitely, whether a particular image belongs in one category or the other.

Diffuse scattering produces images whose intensity increases with thickness up to a certain point depending on the nature of the object and the voltage of the beam, after which the intensity decreases as wide angle scattering predominates.[7] A dark area may therefore indicate the absence of scattering material or it may indicate the presence of a very opaque object. The experimental results show that dark-field images produced by diffuse scattering deteriorate rapidly as the angle of the imaging pencil is increased and are best when taken as close to the limiting aperture as possible, without entering the region of transition between bright- and dark-field. Under the best conditions in this work, resolution in this type of image was in the range 100 to 200A, which is definitely inferior to that in bright-field. Particles smaller than 100A can be detected, but the images are increased in diameter. Factors contributing to the poorer resolution are the greater relative aperture of the scattered beam, the greater effect of spherical aberration for extra-axial pencils, and the greater proportion, at wide angles, of electrons which have suffered velocity loss. Two micrographs, illustrating the comparative quality of a bright-field image and a dark-field image produced by diffuse scattering, are reproduced in Fig. 2. These are micrographs of the well-known tobacco mosaic virus rodlets which are about 150A in width. The rodlets are well resolved in dark-field with some increase in contrast, but the resolution is considerably poorer than in the bright-field micrograph.

Crystalline diffractions, on the other hand, may form relatively sharp images because the beam is usually confined to a narrow pencil which minimizes image broadening by spherical aberration. If an electron microscope specimen contains periodic structure of appropriate dimensions and orientation, coherent scattering will occur, giving rise to well defined pencils, as shown in Fig. 3. The angle 2θ between the incident and diffracted pencils is given (for transmission) by the well-known Bragg law,

$$2d \sin\theta = n\lambda, \qquad (1)$$

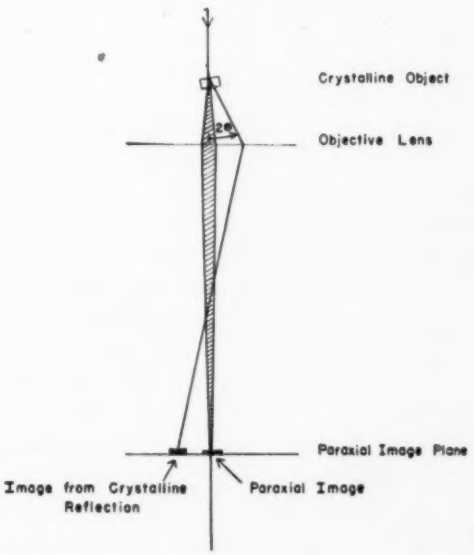


FIG. 3. Electron microscope images of crystalline substances.

* Union Carbide and Carbon, New York.

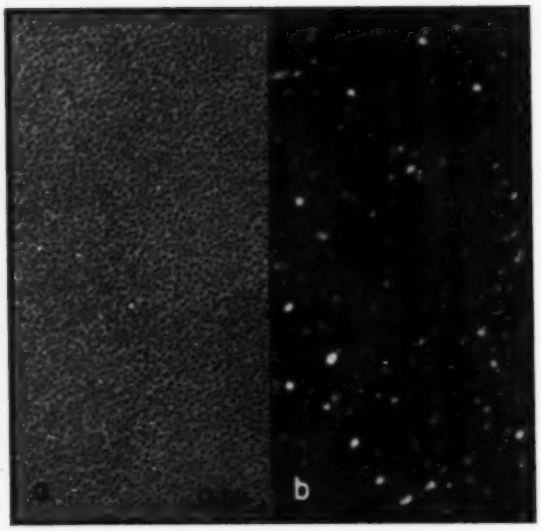


FIG. 4. Evaporated gold film: a, bright-field, b, dark-field.

where d is the interplanar spacing, n is an integer, and λ is the electron wave-length. For the 65,000-volt electrons used in this work, λ is 0.0465A. A spacing of 2A would therefore produce a deflection 2θ equal to about 0.02 radian, which is considerably greater than the optimum relative aperture. As a result of the large spherical aberration, the diffracted rays will be brought to a focus at some distance from the paraxial image, as indicated *in exaggeraged fashion* in Fig. 3. A suitably small diaphragm in the objective would eliminate the diffraction image, but in operation without an aperture, or with a large aperture, faint diffraction images are frequently seen in bright-field. Although the diffracted rays enter the lens at large angle, diffraction images are usually quite sharp because the relative aperture of the pencil is small.

When the objective current is altered, as, for example, in a through-focus series, the diffraction images move with respect to the paraxial image.[17] The reason for the displacement may be seen by referring to the diagram in Fig. 3, where alteration of the objective current alters the deflection produced by the lens. For a given crystal there will be an objective setting for which the diffraction image falls exactly on the paraxial image, but if the Bragg deflection is large compared to the optimum relative aperture of the lens, the paraxial image will not be exactly in focus at this setting because of spherical aberration. In dark-field, diffraction images are usually intense and sharp and may be identified by their relative motion in a through-focus series. In Fig. 4 are reproduced bright- and dark-field micrographs of gold particles produced by vacuum evaporation. The resolution in the dark-field image, which is produced almost entirely by crystalline diffractions, is about 50A. Although the resolution is not quite as good as that in bright-field, it is considerably better than that in images produced by diffuse scattering as typified by Fig. 2.

In accounting for some of the effects observed in electron optical imaging, it is sometimes necessary to allow for the change in electron optical refractive index. For electrostatic fields only, and neglecting relativity, the refractive index for an electron which has been accelerated from rest through a potential V, may be taken as proportional (or equal to) $(V)^{\frac{1}{2}}$. If an electron in a medium of potential V passes to a second medium with a potential $V+\Delta V$, where $\Delta V \ll V$, the index of the second medium with respect to the first is approximately $1+\Delta V/2V$. Thus, if a 60,000-volt electron enters a substance whose inner potential is 12 volts, the refractive index of the substance is 1.0001, which is approximately the magnitude encountered in electron microscopy. The potential inside a solid varies, of course, from point to point, but the assumption of a single average value is frequently useful and permissible. Although the index differs little from unity, the wave-length is also small, so that appreciable phase shifts may occur in traversing electron microscope specimens.

When refraction is included, the Bragg law for reflection (rays incident and emergent at the same plane surface) becomes,[18]

$$\sin^2\theta' = \sin^2\theta - (\mu^2 - 1), \qquad (2)$$

where θ' is one-half the *observed* deflection and μ is the refractive index. The terms on the right-hand side are of the same order of magnitude, and their difference may actually be negative for some spacings in which case reflection cannot

[17] J. Hillier and R. F. Baker, Phys. Rev. **61**, 722 (1942).

[18] G. P. Thomson and W. Cochrane, *Theory and Practice of Electron Diffraction* (MacMillan Company, Ltd., London, 1939).

occur. For true reflection, the observed deflection $2\theta'$ bears no simple relation to the spacing. The index of refraction term may depend on the voltage of the electron beam and the azimuth with respect to the crystal.[19] In any case, the total deflection will be considerably less than would be expected from the simple Bragg law. Where the incident beam makes a very small angle with the surface, as, for example, at the edge of a spherical particle such as the carbon particles to be discussed in Part II, it is to be expected that refraction effects will be appreciable.

Contour phenomena which characterize bright-field images have been ascribed to a type of Fresnel diffraction at edges where the refractive index undergoes an abrupt change.[14] In dark-field, contour phenomena are absent except when the micrograph is taken very close to the edge of the limiting aperture, as shown in Fig. 5 which is a micrograph of carbon particles lying on a film in the region of transition from bright- to dark-field. Edges of particles away from the aperture are bright since the high intensity fringe is deflected under the particles into the opening, and edges toward the aperture are relatively dark since the high intensity fringe is deflected into the mask.

As the resolution limit is approached, refractive index changes appear to be an important factor in determining contrast in the electron microscope image in bright-field. When the object consists principally of particles containing atoms of low atomic number, or when the dimensions of particles approach the resolution limit, the structure is practically transparent to the electron beam. The situation is then very similar to the light microscopic observation of transparent particles that are distinguishable from the surrounding medium mainly through having a different refractive index. A through-focus series of such a specimen, consisting of a 30A evaporated film of CaF_2 on a 20A SiO_2 support, is shown in Fig. 6. Some carbon particles were placed on the film to aid in focusing, and a portion of a carbon particle is included at the bottom of each section. The typical contours at the edge of the large particle show that a is focused above the object plane, b is in focus, and c is focused below the object plane.[14] This series demonstrates remarkable changes in intensity distribution in the image. Regions that are bright in a are dark in c, while in b the structures seen in the two out-of-focus prints are not visible at all. Although there is structure in this micrograph, it is difficult to distinguish with certainty from the grain of the plate. The similarity between intensity changes in this series and intensity changes observed in through-focus observation of transparent particles in the light microscope is evident. Loss in contrast at focus is a serious limitation to bright-field observation. In dark-field, the contrast of crystallites depends on differential scattering rather than total scattering and refraction. In d is a dark-field micrograph of the same specimen showing bright spots originating in small crystals that are not visible in the in-focus micrograph. Although the resolution in dark-field is inferior to that in bright-field, the images in the dark-field micrograph are produced by crystals that are near, or below, the bright-field resolution limit.

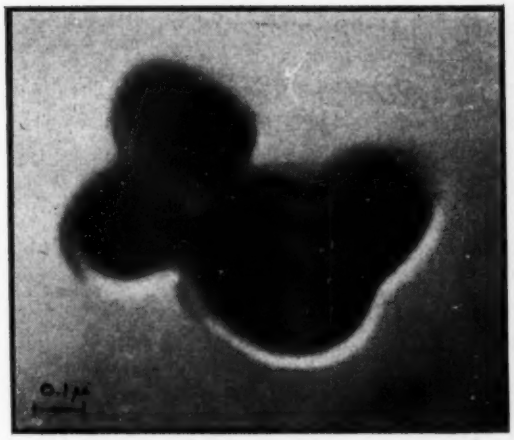


FIG. 5. Carbon particles in region of transition from bright-field to dark-field.

II. Magnesium Oxide

The application of the dark-field method to the study of diffraction phenomena in large, regular-shaped crystals is illustrated by observations made on magnesium oxide. Frequently in bright-field micrographs, faint diffraction images are seen showing anomalous variations in intensity, and these have been studied by Kinder,[20]

[19] C. J. Davisson and L. H. Germer, Proc. Nat. Acad. Sci. **14**, 619 (1928).

[20] E. Kinder, Naturwiss. **31**, 149 (1943).

Heidenreich,[21] and Heidenreich and Sturkey.[22] In dark-field these images are brilliant and well defined, as shown in Fig. 7. In a is a dark-field image of a group of magnesium oxide crystals imaged by (200) reflections and b is an image produced by a (220) reflection. A study of the striated images is of interest since it may provide information regarding the structure of crystals and aid in the interpretation of electron diffraction phenomena. Essentially the fringes are to be ascribed to refractive index variations within the crystal and result from interference between parts of the wave that have traversed the crystal with different optical paths.

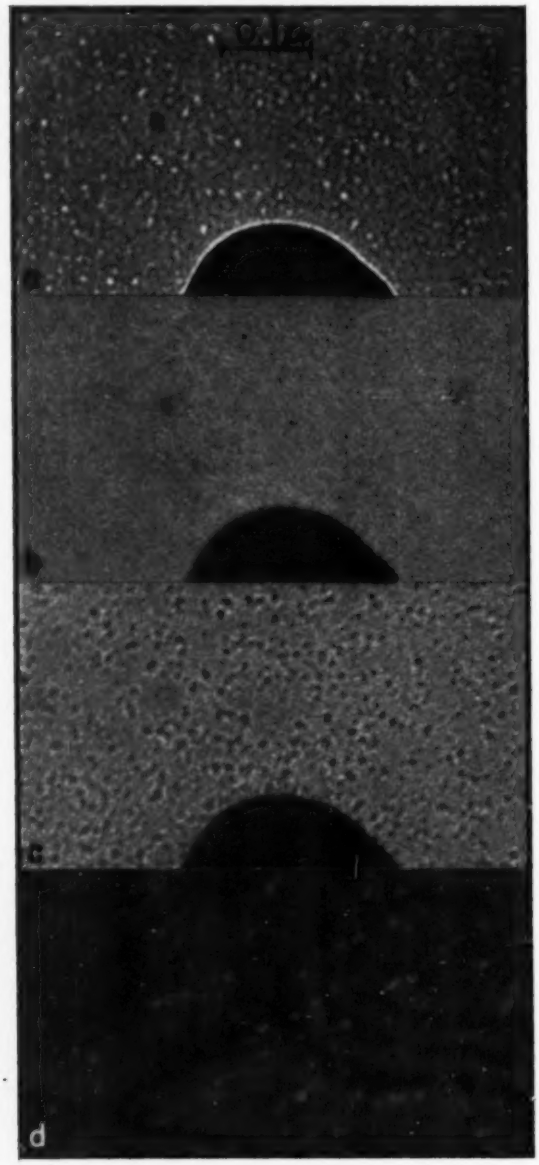


FIG. 6. Calcium fluoride film, average thickness about 30A on quartz supporting film. a, b, and c, are a through focus sequence in bright-field, objective current increasing; d, is dark-field.

Explanations of these phenomena have been discussed by Kossel[23] and Heidenreich and Sturkey.[22] Kossel suggested an explanation using calculations by MacGillivray,[24] whose results were also adopted by Heidenreich and Sturkey. The theory deals with the transmission of electrons through thin lamellae of infinite extent and predicts that the reflected amplitude will vary sinusoidally with the thickness traversed. Dark bands should appear where the thickness is an integral multiple of a critical distance, D_0, which is a function of the accelerating potential and the potential within the crystal. If it is assumed that the theory is applicable to the MgO images, D_0 will be equal to $2s$ for the (200) reflection, where s is the distance between consecutive like fringes in the object plane and $D_0=(6)^{\frac{1}{2}}s$ for the (220) reflection.

Since no direct measurements have been reported of fringe spacing as a function of accelerating potential, measurements were made on the (200) images at two different accelerating potentials (E). According to the theory, D_0 (or s) should be proportional to $(E)^{\frac{1}{2}}$, neglecting relativity. Bright-field micrographs containing several of the (200) reflections were made at two different voltages, and, since the same field was observed, magnification errors are not involved in the comparison of the spacings. If s_1 and s_2 are the spacings for accelerating potentials E_1 and E_2, respectively, we have

$$s_1/s_2=(E_1/E_2)^{\frac{1}{2}}. \tag{3}$$

The voltages were 65,000 and 40,000. Equation (3) then predicts that

$$s_1/s_2=1.28.$$

The average of the measured ratios was 1.35, which differs from the calculated value by about

[21] R. D. Heidenreich, Phys. Rev. **62**, 291 (1942).
[22] R. D. Heidenreich and L. Sturkey, J. App. Phys. **16**, 97 (1945).
[23] W. Kossel, Naturwiss. **31**, 323 (1943).
[24] C. H. MacGillivray, Physica **7**, 334 (1940).

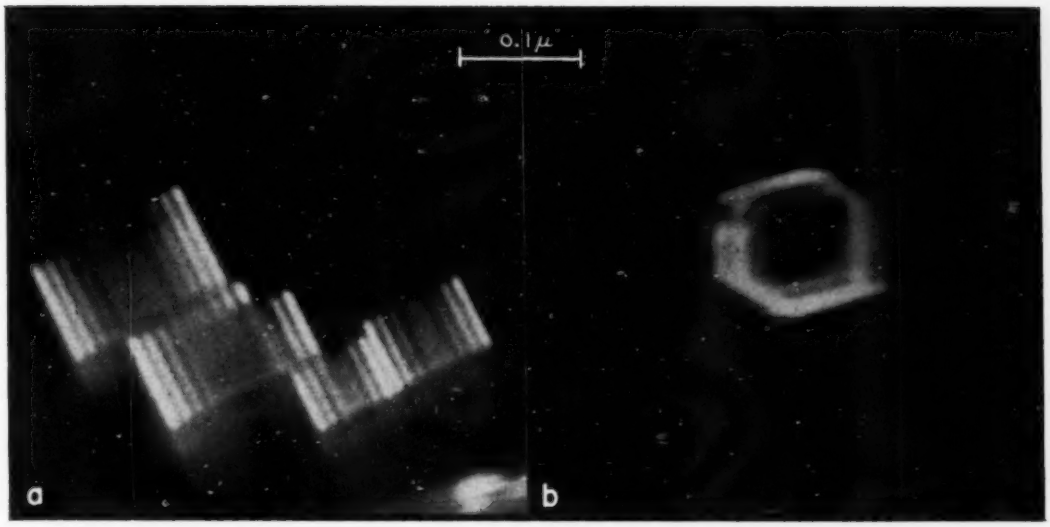


FIG. 7. Dark-field images of magnesium oxide: a, (200) reflection, b, (220) reflection.

6 percent. The measurements then confirm the voltage dependence of s to this order of accuracy.

Measurements of D_0 for about 20 of the (200) images were made from both bright-field and dark-field micrographs taken with 65,000-volts accelerating potential. Values obtained for D_0 were in the range 240 to 340A, with an average of 280A. Although spacings are constant within narrow limits for a given crystal, they may vary from one crystal to another by as much as 10 percent in a single micrograph, where variation cannot be attributed to magnification errors. Also, the variations between micrographs are considered to be significant since magnifications are judged to be accurate within ± 3 percent for these experiments. Kinder[20] reports from observations of bright-field images a range in s from 160 to 200A at 96 kv. Reduced to 65 kv by Eq. (3) and multiplied by 2 to give D_0, these correspond to a range in D_0 from 260 to 320A, which is practically the same as the range found here. Although the measurements reported here are practically the same as those given by Kinder after reduction to the same voltage, they are not in agreement with those reported by Heidenreich and Sturkey.[22] These authors give $D_0 = 420$A at 60 kv and claim that no significant variations occur.

It is interesting to note some recent electron diffraction results on magnesium oxide that are associated with the diffraction images observed in the electron microscope. It has been reported by Hillier and Baker[25] that some of the MgO rings are multiple. Cowley and Rees[26] further report that the multiplicity is explainable on the basis of refraction deflection. They show, in their diagrams, groups of symmetrically placed spots that apparently originate in single crystals and are deflected from their mean position as a result of refraction at the faces of incidence and emergence. These results are consistent with the electron microscope observations in that the spots appear to correspond to the intense reflection images seen in dark-field. Thus for the (220) reflection (oriented as shown in Fig. 7b there are six different refraction deflections. For the (200) reflection (Fig. 7a) two different deflections are possible, but the deflections are tangential to the proper position so that the (200) ring is single and sharp. Similarly, other spots occur which may be interpreted as originating at crystals not only properly oriented for reflection, but at azimuths (orientation about the vector normal to the reflecting planes) similar to those observed for diffraction interference images.

The results suggest that the enhanced intensity of images showing interference phenomena and the occurrence of the interference phenomena at

[25] J. Hillier and R. F. Baker, Phys. Rev. **68**, 56, 209 (1945).
[26] J. M. Cowley and A. L. G. Rees, Nature **158**, 150 (1946).

specific azimuths result from the fact that the electron rays are substantially parallel to a set of crystallographic planes which are normal to the reflecting planes. The variations in D_0 from one image to another might possibly be the result of slight deviations from the precise azimuth with a consequent decrease in the maximum possible optical path differential. At orientations other than these, each ray would presumably pass through a statistical average of the refractive index of the crystal and no systematic interference could occur.

III. Colloidal Particles

An example illustrating the type of information which may be obtained from dark-field observation is shown in Fig. 8. The specimen was made from a red gold sol, about a year old, that had been prepared by reducing gold chloride with yellow phosphorus. In bright-field it is seen to consist of aggregates in which the original separate particles cannot be distinguished. A quite different picture of the structure is obtained in dark-field where the aggregates are imaged by diffusely scattered electrons, and on this diffuse image are super-imposed numerous sharp, bright diffraction images. The dark-field micrograph shows that the aggregates contain discrete crystallites and gives a good measure of their dimensions. Frequently, diffraction images are accompanied by a comet-like tail which may be noted in this and some of the subsequent micrographs. It is not known for certain what causes this anomaly, but since it is usually associated with thick objects it seems to represent electrons that have suffered slight energy loss after they have been scattered coherently.

Micrographs of copper phthalocyanine prepared from a benzene dispersion of the powdered solid are shown in Fig. 9. In a and b are two bright-field micrographs showing typical needle-shaped particles and also granular material. The dark-field micrographs are a little surprising for, as shown in c and d, the needles produce no obvious diffracted intensities, while the granular material produces numerous strong diffraction images. Thus an electron diffraction diagram of this specimen would not originate in the needle-shaped "crystals" as would be expected from the bright-field micrographs, but in the background material. This sample was kindly supplied by the American Cyanamid Company. The example is included to emphasize the value of dark-field electron microscopy in identifying the component in a mixed system producing an electron diffraction diagram.

Other colloidal materials examined included colloidal graphite, zinc oxide, clay, silver, and others. The results were similar to those described above. The dark-field micrographs show

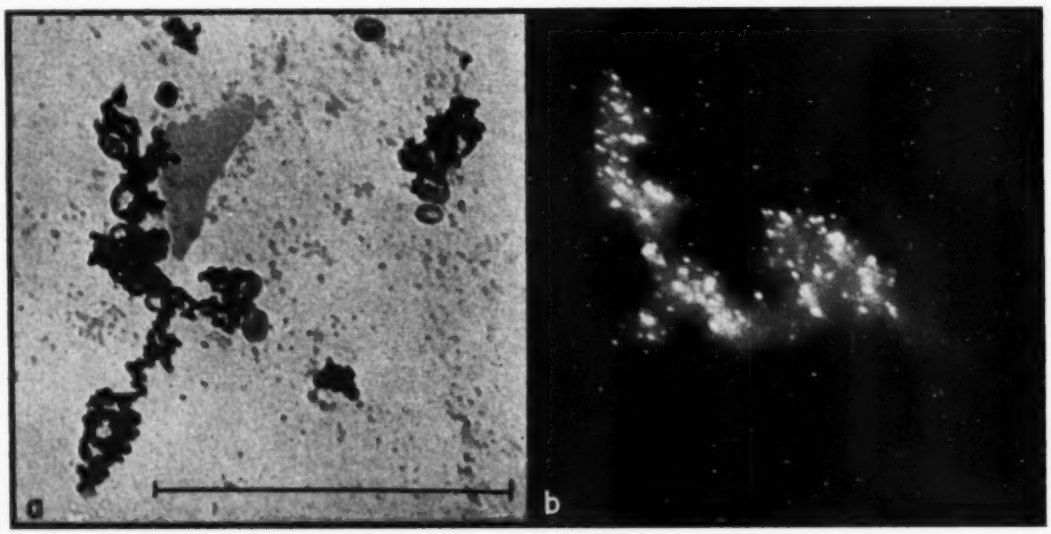


FIG. 8. Aggregates of colloidal gold: a, bright-field, b, dark-field.

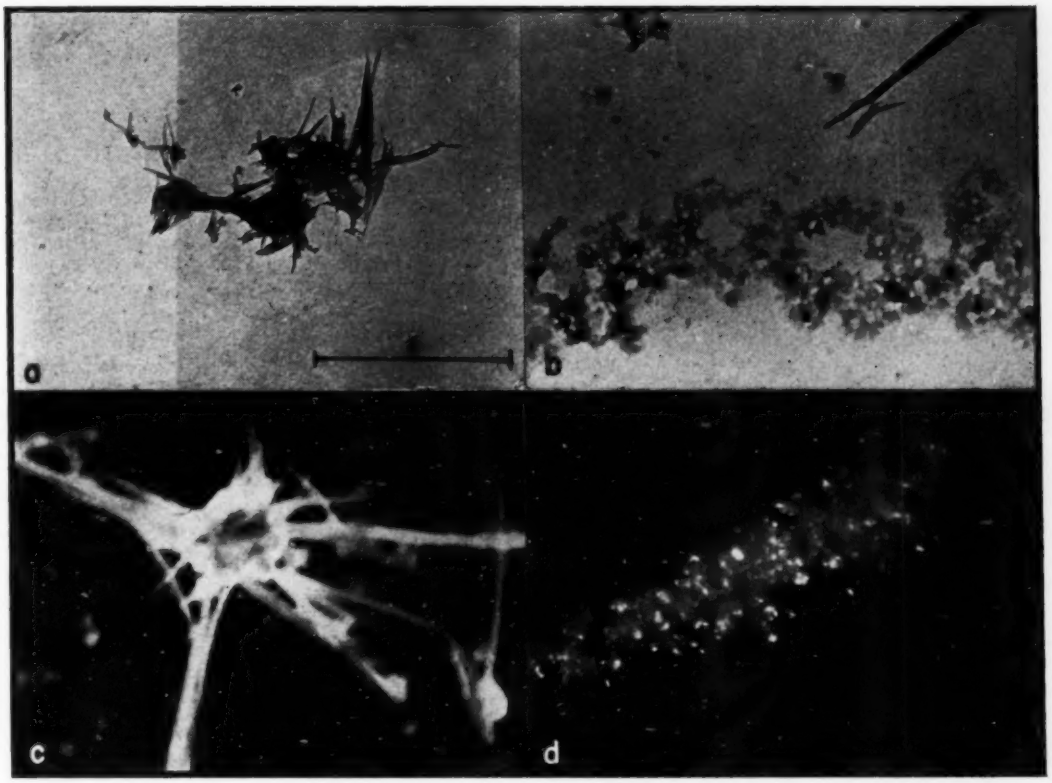


FIG. 9. Copper phthalocyanine: a and b, bright-field, c and d, dark-field.

the size and shape of the structures which would produce the electron diffraction diagram.

IV. Evaporated Films

Vacuum-evaporated films of various metals and compounds were examined in dark-field, the results of which indicate that the dark-field method is particularly useful for studying the shape, size, and distribution of crystalline components in this type of specimen. Most of the films examined contained discrete crystallites whose size and shape varied with the thickness of the film and with the material. An exception is the selenium film shown in Fig. 10. These micrographs show patterns indicating that the ordered regions extend for large distances through the film. Similar patterns have been described in other crystalline lamellae,[27,28] and Boersch[29] has shown that the dark lines in bright-field represent regions where the incident beam is strongly diffracted. A fresh selenium film when first placed in the electron microscope appears uniformly gray, but as the intensity of the beam is increased the dark lines form in a small region and spread throughout the film, presumably as crystallization proceeds. In dark-field, Fig. 10b, bright diffraction streaks appear in a wavy pattern, as would be expected from the bright-field micrograph. Similar patterns were observed from thin films of mica and from gold platelets obtained by very slow reduction of gold chloride.

Thin evaporated films of a number of metals and compounds have been examined in dark-field, some of which were chosen from tables in an electron diffraction study by Germer[30] so that a comparison with electron diffraction results would be available. Typical micrographs of two aluminum films appear in Fig. 11. In a is a bright-field and in b a dark-field micrograph of a 100A aluminum film. The bright-field micrograph shows that the aluminum has formed discrete

[27] M. von Ardenne, Zeits. f. Physik **116**, 736 (1940).
[28] H. Seemann, Naturwiss. **31**, 415 (1943).
[29] H. Boersch, Zeits. f. Physik **118**, 706 (1942).
[30] L. H. Germer, Phys. Rev. **56**, 58 (1939).

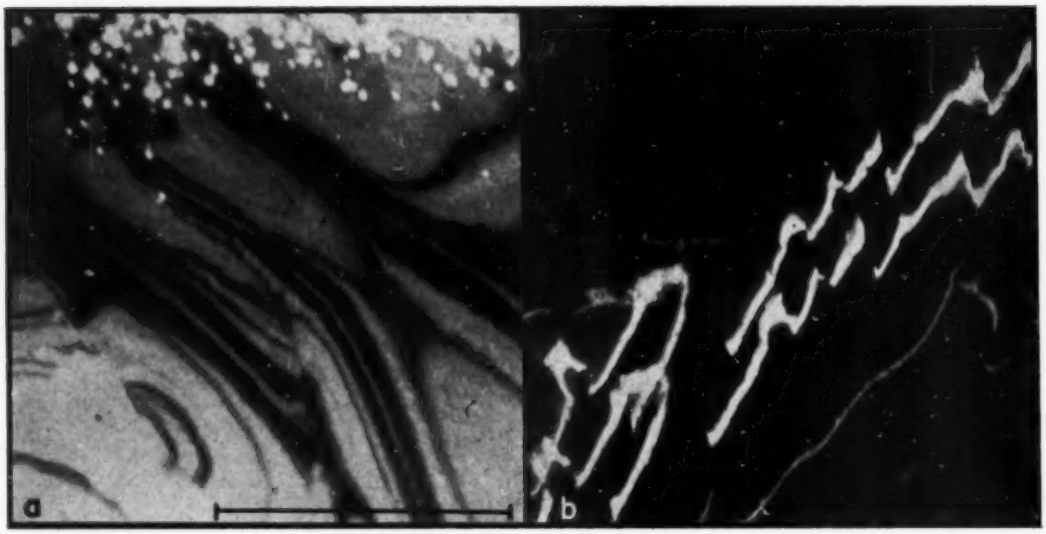


FIG. 10. Evaporated selenium film: a, bright-field, b, dark-field.

particles ranging in size from about 100 to 500A with open channels between them.[31] The dark-field micrograph contains images of from 50 to 500A diameter. The larger dark-field images are well resolved and shaped somewhat like the bright-field images, but there is a much larger proportion of small crystallites than would be expected from the bright-field micrograph alone. The average particle size in bright-field is about 300A, while the average image diameter in dark-field is about 100A. Germer's values for a 108A aluminum film are 59A for the lower limit and 150A probable value. The dark-field measurements are in better agreement with the electron diffraction estimates than are the bright-field measurements, as would be expected. Apparently all the larger particles do not diffract as single crystals.

A thicker aluminum film (about 900A) is shown in Fig. 11b. In this thickness the particles are about 1000A in diameter and show dark and light bands somewhat like those in the selenium film of Fig. 10, but on a smaller scale. Resolution in the dark-field micrograph is relatively poor because of scattering in the thick film, but many images are of the same order of size as those in bright-field and show corresponding variations in intensity.

As an example of a film produced by evaporating a compound, a typical pair of micrographs of a film of antimony trioxide is shown in Fig. 12. Other films such as cuprous chloride (CuCl), calcium fluoride (CaF_2), iron, uranium, silver, gold, and nickel were examined with similar results. Many of these materials had been examined by electron diffraction by Germer, who calculated the crystal dimensions from line breadths. A direct comparison with the electron diffraction results must be approached with some caution, since they were made on different samples, and the character of some of these materials may vary with the conditions of evaporation and subsequent treatment. However, it may be said that the dark-field micrographs are reasonably consistent with the calculations based on line breadths. It is apparent that the location and dimensions of crystalline regions in thin films can be determined by the dark-field method, which has some advantages over the electron diffraction method. However, since the latter also has peculiar advantages, the two methods should be regarded as complementary.

4. RESOLUTION IN DARK-FIELD

A study of resolution in dark-field was made with evaporated films of gold and nickel as test objects. Germer's[30] electron diffraction results show that the crystal size in these films varies with the mean thickness of the film, and this is consistent with both bright-field and dark-field

[31] R. G. Picard and O. S. Duffendack, J. App. Phys. **14**, 291 (1943).

electron microscope observations. For resolution tests it is therefore possible, within limits, to obtain crystals of different average dimensions by evaporating different amounts of material. The bright-field particle diameter appears to be a good measure of the crystal size for gold films. In the case of the nickel films, the particle size was at or below the bright-field resolution limit, and Germer's electron diffraction values were assumed to hold for the similar films used in this work. For example, for a 27A nickel film Germer's probable crystal dimension is 30A, and for a film 14A thick the probable crystal dimension is given as 20A. In bright-field it was possible to make out some structure in films of this thickness, but it was obvious that much of the structure was beyond the capabilities of the instrument. In dark-field, images of discrete crystallites recognizable as such by their characteristic displacement in a through-focus series, could be recorded, but the minimum image diameter was always about 50A. A typical pair of micrographs of a 27A thick nickel film is reproduced in Fig. 13. Diffraction images were also visible in dark-field from a 14A nickel film having probable crystal dimensions of 20A, but the images are not much stronger than the background intensity from the supporting film and the diffuse scattering from the nickel.

As a result of these observations the following conclusions are drawn: crystallites in the range, 15 to 20A, can be detected as relatively large weak spots in dark-field. Crystallites in the range, 30 to 50A, are more sharply imaged, but the apparent image diameter is about 50 to 60A. Crystals greater than about 100A are imaged in dark-field with dimensions approaching the actual crystal size. It is not possible to make a plot of image diameter as a function of crystal size because none of the films consists of uniform particles. However, if such a plot could be made the observations indicate that the image diameter should exhibit a minimum, increasing rapidly toward the small crystals and approaching the true crystal dimensions toward the large crystals.

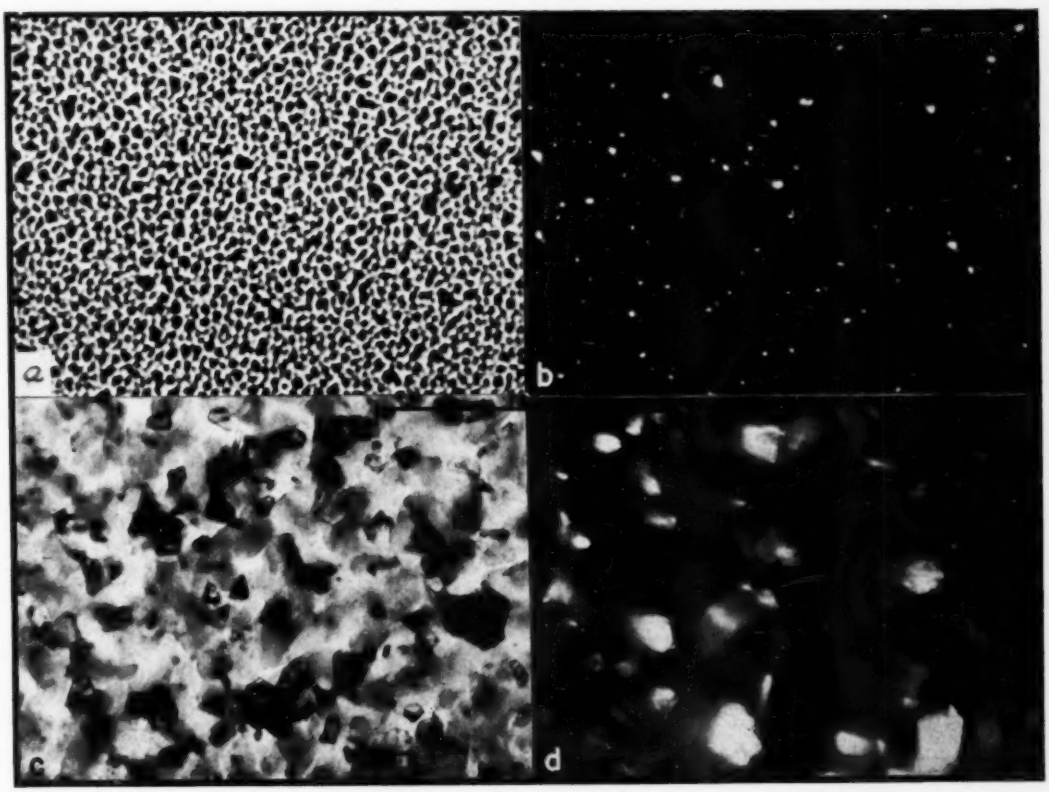


FIG. 11. Evaporated aluminum films: a and b, 100A average thickness, c and d, 900A average thickness; a and c, are bright-field, b and d, dark-field.

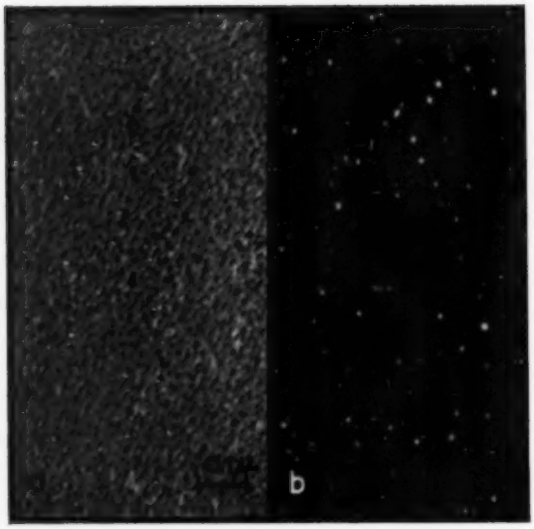


FIG. 12. Antimony trioxide evaporated film: a, bright-field, b, dark-field.

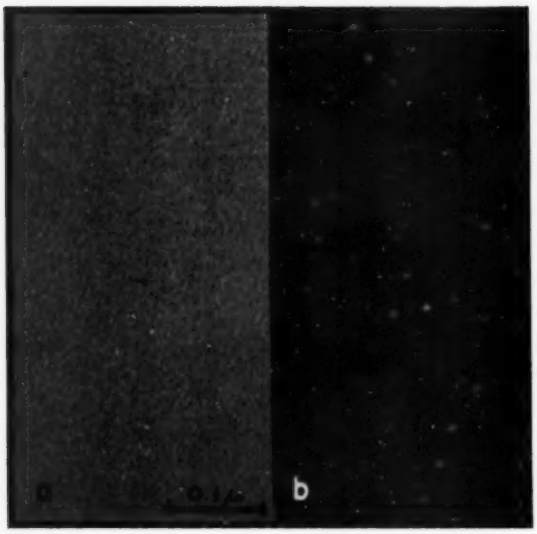


FIG. 13. Evaporated nickel film about 28Å average thickness: a, bright-field, b, dark-field.

It is of interest to consider the possible theoretical basis for the resolution characteristics of the electron microscope for diffraction images in dark-field. Although it cannot be claimed that all the contributing factors are understood at present, it is possible to account for the observations approximately by making some simplifying assumptions and considering the effects of spherical aberration and angular breadth of pencils diffracted by crystallites in the object. The theoretical resolving power is given (from calculations by Rebsch[32]) by the equation,**

$$d=0.4Cf\alpha^3+0.6\lambda/\alpha, \tag{4}$$

where d is the half-intensity breadth of the image, C is a dimensionless constant characteristic of the lens, f is the focal length of the objective, α is the half-angle of illumination (relative aperture), and λ is the electron wave-length.

The first term on the right-hand side gives the image broadening owing to spherical aberration, and the second term is the Abbé formula for the half-intensity breadth of a point image as determined by diffraction alone. It is to be noted that Eq. (4) indicates a rapid deterioration in image quality with increasing α. Also, there is an optimum value of α for which d is a minimum. The optimum value of α and minimum value of d are obtained by differentiating (4) and equating to zero. Substituting for C in terms of d_0, the minimum value of d, (4) becomes

$$d=(0.42d_0^4/\lambda^3)\alpha^3+0.6\lambda/\alpha \tag{5}$$

In dark-field operation the relative aperture of the diffracted pencil is taken as $\frac{1}{2}$ the line breadth at half-intensity. If L is the crystal dimension and λ the wave-length, we set

$$\alpha=\alpha_s+\alpha_c, \tag{6}$$

where $\alpha_s=\frac{1}{2}(\lambda/L)$, i.e., $\frac{1}{2}$ the Scherrer breadth at half-intensity, and $\alpha_c=\frac{1}{2}$ the breadth from some other cause, presumably the angular aperture of illumination.

To be more accurate, resultant line broadening from several causes should be computed as the square root of the sum of the squares of the individual factors, but in view of the uncertainty concerning factors other than Scherrer breadth, the assumption of a more complicated expression for α seems hardly to be justified for the approximate calculations to be made.

For simplification we assume that the diffracted pencil is paraxial, which was not necessarily true for the experimental conditions in this work, but it is an attainable condition where the *incident* rather than the diffracted rays make an angle 2θ with the lens axis. Substituting the expression for α from (6) in Eq. (5), we have

$$d=0.495d_0^4(\tfrac{1}{2}L+\alpha c/\lambda)^3+\frac{0.6}{(\frac{1}{2}L+\alpha c/\lambda)}. \tag{7}$$

[32] R. Rebsch, Ann. d. Physik **31**, 551 (1938).
** See reference 1, p. 699.

This is the expression for the image half-breadth of a "point" in the object. For large particles whose shape is discernible in the image, d would be a measure of the half-breadth of the edges. The formula is only applicable to images of particles whose diameter is less than d. For particles where L is large with respect to d in (7), the image diameter should approach L. As a result of these considerations, curves have been drawn, as shown in Fig. 14, showing the estimated relation between image size (d) in dark-field and crystal dimensions (L). Curve A represents Eq. (7) with d_0 set equal to 30A and $\alpha_c/\lambda = 0.023$ in which α_c is equal to the half-angle of illumination computed from the condenser aperture diameter and the condenser-to-object distance in the electron microscope. The condenser aperture was 0.375 mm in much of the work. At large L, the relation between d and L must approximate the straight line B for which $d = L$. The intermediate range approximated by the curve C represents the image size for small but finite particles. A calculation of this portion presents difficulties, but curves A and B represent limiting conditions, and the transition must be approximately as indicated by C. The curves show a relation which is consistent with the observations. It must be understood that the curves represent approximations (except at large L) and are applicable only to the particular conditions in this study, although the shape should be much the same for all microscopes.

If Eq. (7) holds in general, it should be possible to improve the resolution in dark-field by decreasing d_0, the resolution in bright-field, and by decreasing α_c, the angular aperture of illumination. Although this derivation provides a theoretical relation which is consistent with the observations, it must be regarded for the present as a semi-empirical result, for there may be other factors not considered here which will influence the ultimate resolution in dark-field.

An interesting point in connection with the above computations is worthy of note. If calculations are made for extra-axial pencils at twice the Bragg angle 2θ, it is necessary to have some estimate of the lens constant C in Eq. (4).*** If the resolution in bright-field is 30A and is determined through Eq. (4), then it is readily calculated that $Cf = 65$ cm. Adopting this value for Cf and assuming an extra-axial pencil at an angle corresponding to a 2A spacing with a Scherrer breadth of 5×10^{-4} radian ($L = 100$A), the calculated resolution limit in dark-field is about 1000A. This is inferior by a factor of about 20 to what is actually obtained. With smaller spacings or smaller crystals, the resolution limit is even higher. The reason for this large discrepancy is not entirely clear. In order to obtain approximate agreement between theory and experiment, it is necessary to assume a value of Cf equal to about 2 cm or less. Such a value of Cf is quite reasonable on theoretical grounds, but when introduced into Eq. (4) we should have a resolution of 10A or better in bright-field with appropriate relative aperture. Experimental evidence for resolution at this level is lacking in this work. Thus, in order to account for the observed resolution in dark-field, it is necessary to assume that the optical system is capable of resolving structure considerably finer than has been seen in practice. The failure to observe the required resolution in bright-field may be due to factors not considered in Eq. (4), such as scattering in the object, or it may be due to loss in contrast at true focus, as illustrated in the series of micrographs in Fig. 6.

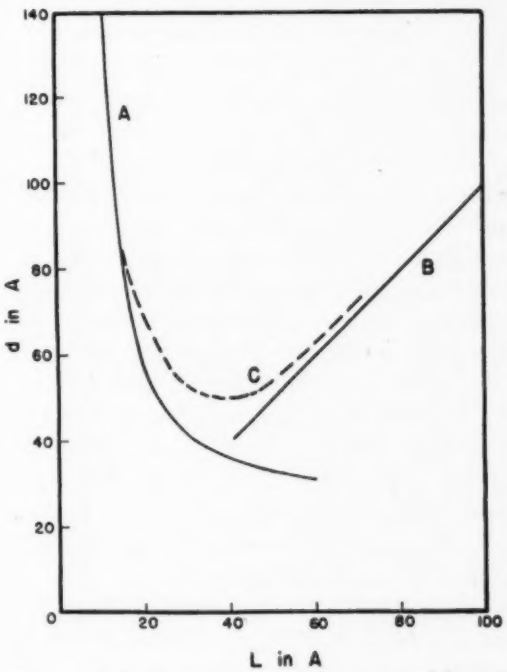


FIG. 14. Estimated image diameter, d, in dark-field for crystals of diameter L.

*** See reference 1, pp. 731–2.

5. DISCUSSION

It is a curious fact that, except in the very early experiments of Boersch, discussions of the dark-field method seem to have been limited to the formation of images by what we have termed diffuse scattering and the essential point seems to have been overlooked, that in studying structure by scattered electrons it is the diffracted component that is more likely to provide interesting information. One has only to consider the practice of electron diffraction to realize that the central diffuse spot is generally of little value, but the diffraction rings or spots are rich in information. The principal value of the dark-field method appears to be in its ability to distinguish and characterize periodic structures producing coherent scattering. Dark-field micrographs of amorphous substances are poor compared to those obtained in bright-field and convey essentially the same sort of information. Furthermore, the fundamental difficulties in confining the diffusely scattered beam to small relative aperture by diaphragms in the objective lens seem to impose permanent limitations on the resolving power attainable by diffuse scattering. Although some gain in contrast might be obtained in some instances, no particularly striking advantage has been demonstrated in this connection.

The dark-field method develops as a natural and useful supplement to the existing methods of bright-field electron microscopy and electron diffraction. Bright-field observations provide a picture of everything in the structure within certain limits set by resolving power and contrast. Electron diffraction provides precise information concerning the periodic structures present, also within limits, but in general at a level of dimensions that cannot be reached by the electron microscope. The dark-field method, supplementing these two, provides information concerning the location, shape, and size of those parts responsible for the electron diffraction diagram. It should be particularly useful in identifying components in mixed systems and should be of value in the interpretation of electron diffraction diagrams, especially where refraction effects are present. A study of the structure of finely divided substances should probably include observations by all three methods.

In view of the fact that the multiple aperture system is admittedly a compromise for convenience, possible refinements in the aperture system might be considered. A more elaborate system fixed to the lens rather than to the object holder would be desirable for recording in rapid succession, dark- and bright-field micrographs of the same part of the specimen. A possible disadvantage of an eccentric aperture is that it transmits only those rays reflected within a certain azimuthal range with respect to the incident beam. An annular aperture would perhaps be desirable or possibly an eccentric aperture that could be rotated about the lens axis. Another characteristic of the system used is that only those crystals are recorded that are in the proper orientation. If it is desired to bring others into reflecting position it would be necessary to provide the means for rotating the object in three dimensions, but this elaboration would present rather formidable mechanical problems. Where numerous crystallites are present in the specimen as has been the case in most of this work, the dark-field observation provides images of a statistical sample of the specimen, which is sufficient for many purposes.

The indices of planes producing diffraction images are not immediately apparent from the electron micrographs, except for such crystals as magnesium oxide where the shape and orientation of the crystals can be observed accurately. In many instances it will be desirable to assign the proper indices to the images. A solution to the problem of indexing might be obtained by measuring the displacement of diffraction images as a function of objective current alterations. Theoretically, the displacements should be related to the Bragg angle if alignment is good and the lens fields are perfectly symmetrical. Before such a correlation could be relied upon, a careful study would have to be made of the effects of lens asymmetries and instrumental alignment. Probably such measurements could not approach the accuracy of measurements from electron diffraction diagrams, but they might provide a useful estimate of the spacings associated with diffraction images.

ACKNOWLEDGMENTS

The author is indebted to Professor G. G. Harvey and to Professor B. E. Warren for assistance given on numerous occasions during the course of this work.

Tensile Properties of Rolled Silver Chloride

B. M. Axilrod and J. J. Lamb
National Bureau of Standards, Washington, D. C.
(Received October 27, 1947)

Tensile tests were made at 25° and 93°C on specimens taken from a rolled silver chloride sheet which was prepared from a single crystal. Tensile strengths of approximately 2500 and 1500 lb./in.^{2} were obtained at 25° and 93°C, respectively; the corresponding elongations at failure were about 15 and 40 percent, respectively. For specimens annealed at 200°C and tested at 25°C, the tensile strength was 75 percent and the elongation about twice that of the corresponding values for unannealed specimens. The stress-strain diagram of the rolled material was determined at 25°C with optical strain gauges while loading and unloading two specimens to maximum strains of 0.0015 and 0.0037, respectively; no effect of this moderate strain on the infra-red transmissivity was observed.

I. INTRODUCTION

SILVER chloride is a material that has become of considerable importance for certain infra-red applications because some of its physical properties, principally the mechanical, are different from those of other halide salts used in infra-red work.[1,2] It melts at about 450°C and can be fused into slabs or ingots that are readily rolled into sheets or films.

Fugassi and McKinney[1] prepared silver chloride films and sheets in thicknesses from 0.01 to 1.2 mm. Their films had good infra-red transmissivity between 1 and 15 microns. They reported that the material is inert to many aqueous solutions and can be cemented readily to glass.

Recently Kremers[2] described the preparation of silver chloride rolled sheet and other optical shapes such as lenses and prisms from single crystals of silver chloride. Kremers indicated that Fugassi and McKinney had produced their sheet from polycrystalline slabs.

Tensile tests on a sample of rolled silver chloride sheet were made at the National Bureau of Standards at the request of the Naval Ordnance Laboratory of the Naval Gun Factory, Washington, D. C. The stress-strain properties were determined at 25°C, and 93°C as received, and at 25°C after annealing at 200°C. The results of these tests are presented here by permission of the Navy Department.

II. MATERIAL

The silver chloride sample was in the form of a sheet about 7 in. square and 0.07 in. thick. It was a commercial grade purchased from the same producer whose material Kremers studied; presumably it was prepared in the manner he described. Chemical and spectrographic analyses were made about 10 months after its receipt to determine the purity of untested portions of the material which had been subjected to storage and some handling. The material was found to be 99.5 percent silver chloride with spectrographic examination showing traces of sodium and iron.

Two small pieces of the silver chloride sheet were annealed, one at 200°C for two hours in a circulating-air oven, the other at about 400°C in a muffle furnace for one-half hour. These pieces and a portion of the original sample were placed between crossed Polaroids and the optical pattern obtained with these combinations observed. The specimen heated at 200°C gave a pattern which was granular with no directional nature in contrast to the original which had a pattern similar to that observed by Kremers.[3] The specimen was lighter in color than the original sample. The specimen heated at 400°C was colorless and gave a pattern with crossed Polaroids which was polycrystalline in nature with square areas about 3 mm on a side.

[1] P. Fugassi and D. S. McKinney, "Preparation of silver chloride films," Rev. Sci. Inst. **13**, 355 (1942).
[2] H. C. Kremers, "Optical silver chloride," J. Opt. Soc. Am. **37**, 337 (1947). Cf. also Chem. Eng. News **25**, 1946 (July 7, 1947).
[3] See reference 2, Fig. 2.

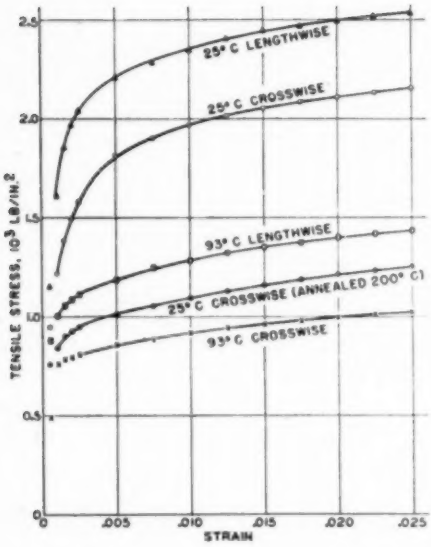


FIG. 1. Tensile stress-strain diagrams for silver chloride sheet determined with Southwark-Peters extensometer. Data at 93°C and for the 25°C crosswise (annealed at 200°C) condition are for one specimen; other data are average for two specimens.

III. TEST PROCEDURES

Tensile Properties

The tensile tests were made in accordance with Method 1011 of Federal Specification L-P-406a[4] except that the lengths of the specimens were between 5.5 and 7 inches. All specimens were 0.75 in. in width with a central portion 0.5 in. wide and 2.2 in. long; the central portion had 3-in. radii at the ends to produce a gradual change in cross section. All tensile tests were made on the 240-lb. range of a Baldwin-Southwark universal hydraulic testing machine of 60,000-lb. capacity. The 240-lb. range was accurate to within 1.2 percent of the indicated load between 10- and 60-lb. load and to within one percent above 60-lb. load.

Eight specimens were tested to failure using a Southwark-Peters non-averaging extensometer,[5] Model PS-6, for measuring strain;[6] the relative rate of cross-head motion as measured during loading was 0.05 in./min. Two lengthwise[7] and two crosswise specimens were tested at 25°C and 50 percent relative humidity after four days conditioning in this same atmosphere. Two crosswise specimens were annealed for 2 hours at 200°C in a circulating-air oven, cooled slowly to room temperature and then tested at 25°C and 50 percent relative humidity. One lengthwise and one crosswise specimen were tested at 93°C and about 2 percent relative humidity after three hours conditioning in the test chamber atmosphere. Tests at 93°C were made with the specimen, tensile grips, and extensometer enclosed in an insulated cabinet having forced air circulation.

To obtain more precise stress-strain data and to get information on the effect of moderate straining on the infra-red transmissivity of the silver chloride sheet, tensile tests were made on two crosswise specimens with Tuckerman optical strain gauges. The strain was read to plus or minus 0.00001 with these gauges, which had a gauge length of 2 in., and a strain range of 0.005. The specimens were strained to a predetermined value and the load released; strain data were taken for both ascending and descending loads to obtain an indication of permanent set. A relative rate of head motion of 0.005 in./min was maintained during loading and unloading.

Infra-red Transmissivity

Infra-red transmissivity measurements for the wave length range 2 to 15 microns were made

TABLE I. Tensile data for specimens of rolled silver chloride sheet at 25°C both before and after annealing at 200°C, and at 93°C.

Specimen orientation[a]	Temp. of test °C	Tensile strength lb./in.²	Stress at 0.1% elongation lb./in.²	Total elongation %
Lengthwise	25	2500	1500	18[b]
Lengthwise	25	2750	1710	9
Crosswise	25	2380	1260	15[b]
Crosswise	25	2500	1140	18[b]
Crosswise (annealed	25	1840	840	>25[b,c]
at 200°C for 2 hr.)	25	1860	—	41[b]
Lengthwise	93	1650	1000	40[b]
Crosswise	93	1350	760	35[b]

[a] The sample was not marked as to rolling direction so lengthwise and crosswise directions were arbitrarily chosen.
[b] Measured with a steel scale after failure; accuracy is about ±2 percent elongation.
[c] Test stopped before complete failure.

[4] Federal Specification L-P-406a, Plastics, organic; general specifications, test methods (January 24, 1944). Government Printing Office, Washington 25, D. C.
[5] Bulletin 162, Baldwin Southwark Division, Baldwin Locomotive Works, Philadelphia, Pennsylvania.
[6] This instrument, with a gauge length of 2 in. and a strain magnification of 400, can record a strain up to 0.11 and hence is suitable for use in the plastic region.
[7] As the direction of rolling on the silver chloride sheet was not known, the lengthwise direction was arbitrarily selected.

within a few hours prior to and subsequent to the tensile tests. These measurements were obtained with a Perkin-Elmer spectrometer having a sodium chloride prism. The tensile specimen was mounted on a holder and was inserted in and out of the beam of radiation in front of the entrance slit of the spectrometer. Tests were made point by point for various wave-lengths in the infra-red region.

IV. RESULTS AND DISCUSSION

Tensile Properties

Stress-strain diagrams for various conditions of test are shown in Figs. 1 and 2. The tensile strength and elongation data are shown in Table I for the several test conditions.

It is of interest to compare these results on the rolled sheet with tensile data obtained by Stepanow[8] on specimens cut from single crystals of silver chloride. Stepanow obtained fracture normally at elongations of 30 to 50 percent at room temperature; he indicated that with careful handling of the specimen much greater elongations could be obtained. Tensile strength values taken from his stress-strain diagrams for two single crystals were approximately 1300 and 1800 lb./in.^{2} at room temperature. A stress-strain diagram for a test made by Stepanow at 200°C indicated a tensile strength of about 570 lb./in.^{2} and an elongation of 40 percent.

The difference in the present tests in the lengthwise and crosswise tensile properties at both 25° and 93°C (Table I and Fig. 1) suggests the presence of anisotropy in the rolled silver chloride sheet. The low tensile strength and the fairly high total elongation obtained on the unannealed specimens indicate that the material has a low resistance to deformation and is ductile.

TABLE II. Infra-red transmissivity of two silver chloride specimens before and after stretching.

Wave-length microns	Maximum strain 0.0015		Maximum strain 0.0037	
	Transmission before stretching %	Transmission after stretching %	Transmission before stretching %	Transmission after stretching %
1.97	65.2	67.0	68.4	67.5
3.00	65.7	—	70.9	68.5
3.99	71.7	70.0	71.4	—
5.04	72.0	—	72.0	71.2
6.04	71.9	—	71.4	—
7.01	72.4	70.2	71.6	—
7.97	74.5	—	71.7	71.9
8.99	73.2	—	71.5	—
9.98	74.2	72.3	72.3	72.0
10.99	75.7	—	73.7	—
12.00	71.5	74.2	72.0	73.0
13.02	74.5	—	71.3	—
14.03	72.4	74.0	70.6	72.5
15.04	72.4	—	73.1	—

[8] A. W. Stepanow, "Die Plastischen Eigenschaften der Silberchlorid- und Natriumchlorid-Einkristalle," Physik. Zeits. Sowjetunion **8**, 25 (1935).

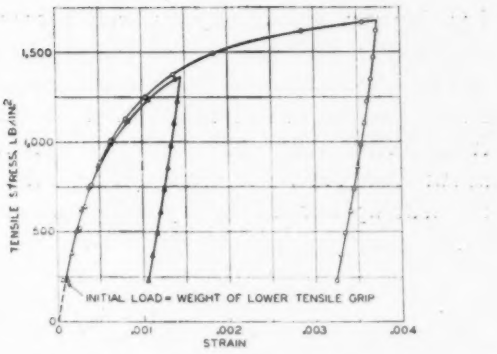


FIG. 2. Tensile stress-strain diagrams for loading and unloading of silver chloride specimens as determined with Tuckerman optical strain gauges.

For crosswise specimens tested at 25°C, annealing them for 2 hours at 200°C results in a 25 percent decrease in the tensile strength and about a 100 percent increase in the elongation at failure, as compared to data for unannealed specimens.

Tensile stress-strain data taken at 25°C with the optical strain gauges on crosswise specimens (Fig. 2) indicate a secant modulus of elasticity at a stress of 750 lb./in.^{2} of the order of 2,000,000 lb./in.^{2}. The curves in Fig. 2 indicate that permanent set occurs at rather small strains. When the material was strained to a value 0.0015, the permanent set was 0.001; for a strain of 0.0037 the permanent set was 0.003.

The question may be raised whether the unloading diagrams (Fig. 2) curve sharply toward the origin below 250 lb./in.^{2}, the stress produced by the lower tensile grip, so that no permanent set exists at zero load. An attempt was made about 10 months after the original test with the specimen which had previously been strained to 0.0015, to settle this question. It appeared that there was no large recovery following straining to 0.0015; the test, however, was inconclusive as the lower grip could not be opened without jarring the strain gauge or bending the specimen.[9]

[9] The testing was done on the 240-lb. range of a 2400-lb. capacity universal hydraulic testing machine which was very similar to the machine used in the earlier tests.

Loading and unloading data were subsequently obtained on the same specimen between 250 and 500 lb./in.^{2} in 50 lb./in.^{2} increments. The stress-strain diagram was linear in this region, the indicated Young's modulus of elasticity value being about 3,000,000 lb./in.^{2}. The agreement between the loading and unloading data was generally within the experimental error.

Infra-Red Transmissivity

The transmissivity data at various regions of the infra-red spectrum of the silver chloride before and after stretching for the two specimens tested with the Tuckerman optical strain gauges (Fig. 2) are given in Table II. The method of measurement of transmission is accurate to ± 1 percent, but other factors make the variation in the values observed before stretching and after stretching somewhat greater. These factors are curvature of the specimen, roughness of the surface, and variation in thickness. When these factors are taken into consideration, it appears that no appreciable change has taken place in the transmissivity of silver chloride due to the stretching.

The values of infra-red transmissivity in Table II are somewhat lower than those reported by Kremers.[2] The discrepancy may be due to the fact that the silver chloride sheet as received was not colorless but had a violet hue; this may have been caused by solarization. Kremers states that optical silver chloride prepared from single crystals is completely transparent and colorless provided extreme care is used in preparing pure silver chloride and that processing of the sheet is carried out in nothing stronger than incandescent light.

The authors wish to acknowledge the cooperation of E. K. Plyler of the Spectroscopy Section of the National Bureau of Standards in obtaining the infra-red transmissivity data, and of the Analytical Chemistry and Spectrochemistry Sections of the same Bureau in analyzing the silver chloride sample.

Calendar of Meetings

February

23–26 Technical Association of the Pulp and Paper Industry, New York, New York
28 American Mathematical Society, New York, New York and Chicago, Illinois

March

1–5 American Society for Testing Materials, Washington, D. C.
2–3 Inter-Society Color Council, New York, New York
4–6 Optical Society of America, New York, New York
22–24 American Chemical Society and the Institute of Technology of the University of Minnesota, Symposium "Modern Instrumental Methods of Analysis," Minneapolis, Minnesota
22–25 Institute of Radio Engineers, New York, New York
26 Institute of Aeronautical Sciences (Flight Propulsion Meeting), Cleveland, Ohio
29–30 Conference on Electron Tubes for Instrumentation and Industrial Use, Philadelphia, Pennsylvania.

April

1–3 American Institute of Electrical Engineers (Great Lakes District Meeting), Des Moines, Iowa
7–9 American Society of Civil Engineers (Spring Meeting), Pittsburgh, Pennsylvania
8–10 Geological Society of America (Cordillera Section), Pasadena California
16–17 American Mathematical Society, New York, New York and Ann Arbor, Michigan
17 American Mathematical Society, Berkeley, California
19–23 American Chemical Society, Chicago, Illinois
21–23 American Geophysical Union, Washington, D. C.
22–24 Acoustical Society of America, Washington, D. C.
22–24 American Philosophical Society, Philadelphia, Pa.
25–30 American Ceramic Society, Inc., Chicago, Illinois
26–28 National Academy of Sciences, Washington, D. C.
28–30 American Institute of Electrical Engineers (Northeastern District Meeting), New Haven, Connecticut
30–May 1 Canadian Institute of Radio Engineers, Toronto, Ontario, Canada

May

9–12 American Institute of Chemical Engineers, Cleveland, Ohio
17–21 Society of Motion Picture Engineers, Santa Monica, California
30–June 2 American Society of Refrigerating Engineers, Swampscott, Massachusetts
30–June 4 American Society of Mechanical Engineers (Semi-Annual Meeting), Milwaukee, Wisconsin

Letters to the Editor

Note on the Anomalous Physical Properties of Processed Atmospheric Nitrogen

CLARKE C. MINTER
Washington, D. C.
October 31, 1947

ABOUT twelve years ago it was observed by the writer that in order to produce a gaseous mixture having a definite thermal conductivity (about 95 percent that of air), less CO_2 was required in the CO_2N_2 mixtures than in the CO_2-air mixtures. CO_2 was added to N_2 from cylinders, or to dry air, to produce a mixture having a given conductivity, the composition being determined by analysis with an Orsat apparatus. In every case it was found that the CO_2 content of the mixture containing nitrogen was only about 75 percent that of the mixture containing air, whereas it would be expected from the relative conductivities of air and atmospheric nitrogen that the CO_2 content of the mixture containing nitrogen would be approximately the same as that of an air mixture having the same thermal conductivity.

When this abnormal thermal conductivity was first observed it was thought to be due to impurities in the processed nitrogen. At that time this question was not investigated; and the use of cylinder nitrogen for making up the gas mixtures was discontinued. The same result has been observed from time to time in the last few years with every cylinder of processed nitrogen used, and it was finally decided to check some of the other physical properties of the nitrogen, and to run down the cause of the anomalous behavior.

The producers of the nitrogen (made from liquid air) state that the argon content was less than 0.05 percent, and that there was no other impurity which could produce such an abnormal result. It was definitely established by the writer that the processed nitrogen used in these tests contained no other impurity having a lower thermal conductivity. It was found that the nitrogen contained no hydrocarbon, no nitrous oxide, or any other impurity which would have to be present in the quantities necessary to reduce the thermal conductivity to the extent observed. It was also observed that the thermal conductivity at room temperature of the cylinder nitrogen was not changed by passing it over copper or copper oxide heated to 600–700°C.

TABLE I. Comparison of the viscosities of dry air (μ air) and dry processed nitrogen (μN_2) at 25°C.

Processed N_2 ml/min.	Air ml/min.	Ratio $^{\mu}N_2/\mu_{air}$	μ Pure N_2	μ Air	Ratio $^{\mu}N_2/\mu_{air}$
134.2	131	0.976	173.8*	182.0*	0.954
166.0	161	0.970			
		0.973			

* Calculated by means of Sutherland's formula from handbook values for 0°C.

TABLE II. Specific gravity of processed nitrogen relative to dry air as measured with the Edwards gas density balance. (Dry air = 1.)

Processed nitrogen	Pure nitrogen	Atmospheric nitrogen
0.985	0.9672	0.9721**
0.984		
0.975		
0.9877		
0.982		
0.9827		

** From the handbooks.

It was found experimentally that at room temperature the processed nitrogen had the same thermal conductivity as dry air containing 3.6 percent CO_2. Atmospheric nitrogen free from argon should have a thermal conductivity about the same or slightly greater than that of air. Hence it should be necessary to add a small quantity of carbon dioxide to such nitrogen to reduce its conductivity to that of air. Instead it is found that an appreciable quantity of carbon dioxide has to be added to air to reduce its conductivity to that of processed nitrogen containing only a trace of argon. It was felt that since the thermal conductivity of the processed nitrogen appears so abnormally low, then perhaps its viscosity and specific gravity would also be abnormal. These expectations have been confirmed experimentally.

The viscosity of the processed nitrogen was compared with that of dry air by measuring the volumes flowing through a given capillary under a given pressure head in unit time, the viscosity being inversely proportional to the volume measured. The specific gravity of the same sample of processed nitrogen was compared with that of dry air by means of the Edwards gas density balance. The results of these tests are given in the Tables I and II.

A sample of the nitrogen was analyzed in a mass spectrometer, but there was no evidence of the presence of N_3 or N_4. It is now planned to investigate the properties of pure nitrogen which has been liquified and evaporated.

Self-Diffusion in Iron

C. ERNEST BIRCHENALL AND ROBERT F. MEHL
Metals Research Laboratory, Carnegie Institute of Technology, Pittsburgh, Pennsylvania
December 1, 1947

THE rates of self-diffusion in alpha- and gamma-iron have been determined by the use of radioactive tracer techniques similar to those employed by Steigman, Shockley, and Nix.[1] The radioactive iron was supplied by

TABLE I. Comparison of frequency factors from experiment and Dushman-Langmuir equation.

	Experimental	$D-L$
$A\alpha$	3.4×10^{4}	0.51
$A\gamma$	1.0×10^{-4}	0.32

the U. S. Atomic Energy Commission. The iron was the Westinghouse spectroscopic standard, Puron.

Measurements covered the temperature range 715 to 887°C in the alpha-phase and 935 to 1112°C in the gamma-phase. When both curves are extrapolated to the transformation temperature, 910°C, it is found that the rate of self-diffusion in alpha-iron is about 100 times as great as the rate in gamma-iron. This in itself is not surprising since J. L. Ham[2] found molybdenum to diffuse 80 to 90 times as fast in alpha-iron as in the gamma-phase at the same temperatures. However, unlike Ham, we observe a great difference in the activation energies in the two phases with correspondingly larger differences in the frequency factor.

The equations giving the best agreement with the six measured points in the alpha-phase and the eight points in the gamma-phase are

$$D_{\alpha Fe} = 34000e^{-77200/RT},$$
$$D_{\gamma Fe} = 0.00104e^{-48000/RT}.$$

The points have been corrected for the change in diffusion distances introduced by cooling the samples to room temperature for measurement. These corrections are between 2 and 3 percent of the value of D.

Substituting the measured activation energies, Q, and the interatomic distances, d, for iron into the Dushman-Langmiur equation permitted independent calculations of the frequency factor, A. A is given by the expression $A=(Q/Nh)d^2$, where N is Avogadro's number and h is Planck's constant. A comparison (Table I) of the experimental and calculated values shows that the agreement is very unsatisfactory and that this equation will not describe the experimental data.

[1] J. Steigman, W. Shockley, and F. C. Nix, Phys. Rev. **56**, 13 (1939).
[2] J. L. Ham, A.S.M. **35**, 331 (1945).

Temperature Gradients for Convection in Well Models

MARY JANE AULD
University of North Carolina, Chapel Hill, North Carolina
November 14, 1947

PRELIMINARY experimental data have been obtained in small-scale tests in an attempt to check the theoretical calculations of A. L. Hales[1] relative to minimum temperature gradients required for the initiation of convection currents in water-filled wells. Half-meter long glass tubes of 5-, 7-, and 9-mm inside diameters served as "wells;" the bottoms were sealed off, and iron-constantan thermocouples were inserted in wall notches at approximately 12-cm intervals to provide temperature measurements. Each tube was placed in a vertical position, filled with water, and heated at its lower end by a steam bath.

Temperatures and elapsed times were recorded, heat being applied at the time of first recording; and the tubes were constantly watched for signs of convection. Convection was considered observable at the first signs of motion of the inevitable small particles of impurities suspended in the water. As observed, initial convection was definitely localized, i.e., one first observed the motion of particles at the first thermocouple junction above the steam, but could observe no motion 5 cm above; motion was observed at the second thermocouple some time later. From these data, isochronal curves were plotted for 5-minute intervals, showing water temperatures *vs.* distance above steam bath; hence, both local and average-over-depth gradients could readily be determined for the instant of onset of convection.

Columns 1, 2, and 3 of Table I summarize these observations; in column 3, the depth refers to length of column of liquid above the steam bath, generally about 40 cm. Column 4 shows Hales' calculated gradients, which correspond to the radii used in these tests. Although Hales' calculated gradients are systematically less than the observed values, there is a qualitative agreement; indeed, it appears that the observed gradients can be represented as proportional to r^{-4} much better than to r^{-3} and r^{-5}. The significance of the quantitative disagreements shown by Table I is not clear at the moment.

The author is indebted to Mr. C. W. Horton, of the University of Texas, and to Dr. F. T. Rogers, Jr., for suggesting this investigation and for valuable discussions during its course.

[1] A. L. Hales, Mon. Not. Roy. Astron. Soc. Geophys. Supplement **4**, 122 (1937).

TABLE I.

Tube radius r(cm)	Temperature gradient (°/cm) at onset of convection: Local	Average-over-depth	Hales' $0.0014/r^4$
0.25	1.5	0.5	0.35
0.35	0.29	0.23	0.093
0.45	0.17	0.085	0.034

New Books

X-Rays in Practice

BY WAYNE T. SPROULL. Pp. 615, McGraw-Hill Book Company, Inc., New York, 1946. Price $6.00.

The jacket of this book declares that the contents furnish "a readable and comprehensive treatment, giving the reader a broad understanding of x-rays, their nature, and the many purposes for which they may be used." The reviewer finds this an accurate description.

This book is not intended as a treatise on any specific field in which x-rays are employed, but, rather, seeks primarily to provide the desirable background for understanding such fields, and secondarily to give introductions to each field. Of the 24 chapters, the first ten are devoted to the fundamental characteristics of x-rays and the equipment for generating them. This is followed by three chapters describing non-diffraction uses, chiefly in the medical and industrial radiology fields. Ten more chapters are concerned, directly or indirectly, with x-ray diffraction

and its application. Crystallography and crystal structure investigation are the main theme of five of these ten chapters, the remaining five comprising applications of x-ray diffraction technique. There is a final chapter on the related field of electron diffraction and its application.

The reviewer can recommend the first part of the book. To one who wishes to make actual use of x-rays in one of its several fields of application, this part offers a basic but comprehensive account of x-rays, their theory, and how to generate them. Since the style is very readable, this should be especially appreciated by the non-physicist whose curriculum of study has not included a formal course in the physics of x-rays. These remarks should not be construed to mean that the account is popular, for it is technical and sound, but not stilted.

The reviewer is less enthusiastic about the five chapters concerned with crystallography and the fundamentals of x-ray diffraction. This contains considerable material which is out-of-date, and, as a consequence, hardly presents a present day perspective in a science which has made tremendous strides in less than a generation. In Chapter 16, the Bragg x-ray spectrometer is discussed, along with the original Bragg method of deducing the structures of "rocksalt" and "sylvine." Structure factors are illustrated only for the degenerate cases of NaCl and ZnS. It is stated, with a 1931 footnote as authority, that the Laue method of crystal analysis is still used today by some of the foremost crystal analysts. Zwicky is cited as an authority on the nature of imperfection in crystals, with a 1934 reference. The remark is made that the only system of space group nomenclature that nearly everybody understands is the Schoenflies notation. All this represents the x-ray analysis of the 1930's and earlier, and is inappropriate in a book published in 1946.

In spite of this adverse criticism of part of the book, the reviewer feels that the rest of the book fulfills a need so well that the book is a worthwhile contribution to the literature. This part of the book constitutes a valuable introduction to the properties and devices for the generation of x-radiation, and it is replete with helpful and interesting suggestions on the techniques concerned with x-ray production.

M. J. Buerger,
Massachusetts Institute of Technology

New Booklets

X-Ray Division of Eastman Kodak Company, Rochester 4, New York, is offering free a new conversion chart for x-ray diffractionists which permits rapid calculation of exposures with films of varying characteristics of speed, contrast, and graininess.

Ward Leonard Electric Company, 55 West Jackson Boulevard, Chicago 4, Illinois, announces publication of its new catalog D-30, which fully describes and illustrates a line of stock units in resistors, rheostats, and radio amateur relays. Address request to the Radio and Electronic Distributor Division at the address shown above.

Tube Department of RCA Victor Division, Harrison, New Jersey, has issued a new edition of the *RCA Receiving Tube Manual*, RC-15. This manual presents many new features, the result of rapid progress in electronics during the war years. 256 pages, 35 cents. Available at the address given above or from RCA receiving tube distributors.

In response to a strong demand for technical information on printing electronic circuits, the National Bureau of Standards has published NBS Circular 468, *Printed Circuit Techniques*. 43 large pages, 39 figures. Available from the Superintendent of Documents, U. S. Government Printing Office, Washington 25, D. C., at 25 cents per copy.

Here and There

New Appointments

R. S. Neblett has been named administrator of the Nucleonics Project of General Electric Company, which consists of work on atomic energy for the U. S. Atomic Energy Commission.

Henry C. Froula, recently Secretary of the Physics Club of Chicago and electron microscopist at Armour Research Foundation, has been appointed to the physics staff of Los Angeles City College.

H. Hugh Willis has been appointed Director of Research and Development for the Kellex Corporation atomic energy subsidiary of the M. W. Kellogg Company.

William G. Pollard has been named Executive Director of the Institute of Nuclear Studies, Oak Ridge, Tennessee. Dr. Pollard has been acting director of the Institute since February 1947.

Awards

According to Science, The Research Corporation recently made two awards, each consisting of a plaque and a $2500 honorarium, to Lee A. DuBridge, President, California Institute of Technology, "for his outstanding scientific achievements in directing the Radiation Laboratory of OSRD in the field of microwave radar research, development, and application to national defense," and to Merle A. Tuve of the Carnegie Institute of Washington, in recognition of his scientific contributions in making possible the proximity fuse and his administration of the various groups which equipped the Armed Forces with the device.

Science also reports that Robert Jemison van de Graaff, Massachusetts Institute of Technology, has been awarded the 24th Duddell Medal of the Physical Society, London, in recognition of the invention and development of his